Nonparametric Identification of the Distribution of Random Coefficients in Binary Response Static Games of Complete Information

Fabian Dunker
University of Canterbury*

Stefan Hoderlein
Boston College†

Hiroaki Kaido
Boston University‡

Robert Sherman
California Institute of Technology§

June 5, 2017

Abstract

This paper studies binary response static games of complete information allowing complex heterogeneity through a random coefficients specification. The main result of the paper establishes nonparametric point identification of the joint density of all random coefficients except those on interaction effects. Under additional independence assumptions, we identify the joint density of the interaction coefficients as well. Moreover, we prove that in the presence of covariates that are common to both players, the player-specific coefficient densities are identified, while the joint density of all random coefficients is not point identified. However, we do provide bounds on counterfactual probabilities that involve this joint density.

Keywords: Binary Response, Static Games, Complete Information, Heterogeneity, Nonparametric, Identification, Random Coefficients.

*Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand, Email: fabian.dunker@canterbury.ac.nz.

†Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, Email: stefan_hoderlein@yahoo.com.

‡Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA, Email: hkaido@bu.edu.

§Division of Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA, Email: sherman@amdg.caltech.edu.

This paper contains parts of the identification section of Dunker, Hoderlein, and Kaido (2014) "Random Coefficients in Static Games of Complete Information,"(DHK, henceforth). The estimation part of DHK will be part of a companion paper. We thank Andres Aradillas-Lopez, Arie Beresteanu, Iván Fernández-Val, Jeremy Fox, Eric Gautier, Yuichi Kitamura, Elie Tamer, Whitney Newey, seminar participants at Boston College, Boston University, Harvard, UCL, University of Pittsburgh, Yale, and conference participants at the Second CIREQ-CEMMAP Workshop on Incomplete Models and the 2013 North American Winter Meeting of the Econometric Society in San Diego. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft in the SFB 823 (projects C1 and C4) and by the Ministry of Education and Cultural Affairs of Lower Saxony in the project "Reducing Poverty Risk"(Dunker) and National Science Foundation SES-1357643 (Kaido).

1 Introduction

Motivation. Unobserved heterogeneity across cross sectional units is ubiquitous in situations of strategic interaction. The decisions of airlines to enter a given local market, for instance, may be dramatically influenced by unobservable factors not captured by observables like market size or average income. Similarly, there may be profound differences in the work and retirement decisions of married couples that are not sufficiently explained by observed variables like age, number of children, or religious background. Yet, understanding the extent of these differences across many local markets or many couples is crucially important for making effective policy decisions.

In this paper, we adopt a random coefficients approach to model such heterogeneity of players across different cross sectional units, e.g. local markets or couples. We consider the most basic model of strategic interaction in a binary response, two player one-shot complete (perfect) information game. To this end, we study a binary response linear index dummy endogenous variable simultaneous equation model. This model has been extensively analyzed with non-random coefficients and a scalar unobservable, see Amemiya (1974), Heckman (1978), Bjorn and Vuong (1985), Bresnahan and Reiss (1990, 1991), Berry (1992), and Tamer (2003). More recently, this line of work has been extended by Kline (2015) to allow for a scalar heterogeneous random parameter on the interaction term. Fox and Lazzati (2015) consider a complete information game with multiple players and study its relation to the demand of bundles, while allowing for unobservable heterogeneity as in Kline (2015). In contrast to all these references, we focus on the two player game with possibly high dimensional unobservable heterogeneity.

To formalize our approach, under suitable assumptions on the solution of the game our structural model maps into the following reduced form system of equations:

$$\begin{aligned} Y_1 &= \mathbf{1}\{(B_1 + \Delta_1 Y_2)X_1 - Z_1 < 0\} \\ Y_2 &= \mathbf{1}\{(B_2 + \Delta_2 Y_1)X_2 - Z_2 < 0\}. \end{aligned} \tag{1.1}$$

Here Y_j denotes a binary action that player j may take. In the above retirement example, $Y_1 = 1$ denotes the decision of spouse 1 to retire. In the market entry example, $Y_2 = 0$ denotes the decision of firm 2 not to enter the market. We assume that this decision, for each player $j = 1, 2$, is determined by whether the latent utility $Y_j^* = Z_j - (B_j + \Delta_j Y_{-j})X_j$ is above or below a threshold normalized to be zero; if the utility is above the zero threshold, (equivalently, if $(B_j + \Delta_j Y_{-j})X_j - Z_j < 0$), player j chooses $Y_j = 1$. This utility is partially determined by observables, specifically, the covariates $\tilde{X}_j = (X_j, Z_j)$, where X_j includes a constant and Z_j is a scalar[1], as well as the action of the other player, Y_{-j}. However, it is also partially determined

[1]We distinguish notationally between a single covariate Z_j, whose coefficient we normalize to be unity almost surely, and the remaining vector of covariates X_j. As is implicit in the binary nature of the actions, because of the indicator function there is a choice of normalization. Throughout this paper, we assume that the sign

by the unobservables B_j and Δ_j, which determine how observable covariates, as well as the action of the other player influence the latent utility.

The key innovation in this paper is allowing all the variables, including the unobserved parameters, to vary across the population, thus adopting a perspective of extensive heterogeneity. We then provide a framework in which we establish point identification of the nonparametric distribution of the random parameters. Our main identifying assumption is that all the observable covariates $(X, Z) = (X_1, X_2, Z_1, Z_2)$ are independent of the unobserved random vector $(B, \Delta) = (B_1, B_2, \Delta_1, \Delta_2)$, conditional on additional covariates W. We think of the system (1.1) as a system of simultaneous equations and note, as is well known in the literature, that the properties of the model change fundamentally with the sign of the interaction effects, see, e.g., Bresnahan and Reiss (1991), or Tamer (2003). Therefore, we focus largely on subcases. In particular, we start out with the case where, in every market we use for identification, the players behave as "strategic substitutes", which is central to the literature on market entry. In our setup, this means that there is always a negative externality from a player entering the market on the net utility of the other player, but to a varying degree across markets. We also cover the case of "strategic complements", where the other player's action positively affects the player's own utility, which is plausible, for example, in the joint work and retirement decision.

Our main result states that, in the case of strategic substitutes, the joint densities of $B + \Delta$ and B, respectively, are point identified. The intuition behind this result is as follows. One can relate the joint characteristic function of B to the conditional entry probability of $(Y_1, Y_2) = (0, 0)$ and similarly that of $B + \Delta$ to the conditional probability of $(Y_1, Y_2) = (1, 1)$. The exogenous variation of the covariates (Z, X) then allow us to trace out the joint characteristic functions. To this end, the variation in the covariates has to be sufficiently large. Specifically, we require that the covariates $Z = (Z_1, Z_2)$ have joint full support. The support requirement on other covariates X depends on assumptions we are willing to place on the density of interest. If we assume that the characteristic function of the density is analytic, it suffices for X's support to contain a small open set. This requirement can be met in various applied examples, and hence we view it as the main empirically relevant requirement. Alternatively, if we assume that the density is entirely unrestricted, the covariates must have full support. While it highlights the required variation in the covariates for achieving point identification without any restriction on the density, this requirement is frequently unrealistic and also has limitations in terms of consistency with other assumptions. Once the characteristic functions of B and $B + \Delta$ are recovered, we may identify the density of Δ via deconvolution under the additional assumption that B and Δ are independent.

In either case, this result implies that the joint density of the interaction effects, f_Δ, is only

of one of the original random coefficients be the same for the entire population. This allows us to normalize by this random coefficient and it is the corresponding variable which we denote by Z_j. In an earlier version of the paper (DHK (2014)), we show that the model is actually identified by (only) imposing a dual hemisphere normalization condition, but the economic benefits of this greater generality are minor.

set identified in general. However, under additional independence assumptions we obtain point identification of f_Δ and $f_{B\Delta}$ as well. In the case of strategic complements, we show that the joint characteristic function of $(B_1 + \Delta_1, B_2)$ can be related to the conditional probability of $(Y_1, Y_2) = (1, 0)$ and similarly that of $(B_1, B_2 + \Delta_2)$ is related to the conditional probability of $(Y_1, Y_2) = (0, 1)$. Therefore, these entry outcomes provide information that allows recovery of the marginal distribution of the interaction effects under additional independence assumptions. For example, Δ_1's distribution can be recovered from those of $B_1 + \Delta_1$ and B_1, however, the joint density of interaction effects f_Δ remains only partially identified in this scenario.

The identification principle put forward is constructive and can be employed to construct nonparametric sample counterpart estimators, whose analysis we defer to the companion paper. In addition to contributing to the abstract understanding of these models, we also augment our approach to cover issues that are important in applications. In particular, as part of the main result we consider the case where some of the covariates are market-specific and/or discrete, and show that important features such as the mean and the marginal distributions of the random coefficients can be point identified.

We further investigate the consequences of the presence of market-specific variables and formally show that their presence generally prohibits us from identifying the joint distribution of *all* coefficients. However, we also show that the joint densities of some subvectors of the random coefficients are still identified, allowing recovery of economically meaningful objects such as the distribution of the effect of a player-specific variable (e.g. market presence) through strategic interactions. Using this result, we further provide partial identification results for counterfactual objects that depend on the joint distribution of the coefficients.

Contributions relative to the Literature. Simultaneous discrete response models have been studied extensively. Much of the literature has focused on identification and estimation of structural parameters that are assumed to be fixed across markets. Ciliberto and Tamer (2009), for example, estimate an entry model of airline markets assuming that the parameters in the airlines' profit functions are either fixed or depend only on observable characteristics of the markets. A novel feature of our model is that the structural parameter may vary across markets following a distribution which is only assumed to satisfy mild assumptions.

A key challenge for the econometric analysis of this class of models is the presence of a region in which each value of payoff relevant variables may correspond to multiple outcomes. Tamer (2003) calls such a region the *region of incompleteness*. Early work in the literature including Amemiya (1974), Heckman (1978), and Bjorn and Vuong (1985) assume that a unique outcome is selected with a fixed probability. More recently, Bresnahan and Reiss (1990, 1991) and Tamer (2003) show that structural parameters can be identified without making such an assumption. The former treats the multiple outcome as a single event and identifies the structural parameters by analyzing the likelihood function. The latter treats the multiple outcome as is, but requires the existence of special covariates that are continuously distributed with full support. See also

Berry and Tamer (2006) for extensions.

As already mentioned, we nonparametrically identify the distribution of random coefficients without making any assumption on the equilibrium selection mechanism, but utilize the assumption that covariates are continuously distributed. Other recent work on identification in complete information games (with fixed coefficients) includes Bajari, Hong, and Ryan (2010), who establish identification of model primitives including an equilibrium selection mechanism using exclusion restrictions, Beresteanu, Molchanov, and Molinari (2011) and Chesher and Rosen (2012), who apply the theory of random sets to characterize the sharp identification region of structural parameters, and Kline and Tamer (2012), who derive sharp bounds on best response functions without parametric assumptions. Less closely related is the work on identification, estimation and testing in games of incomplete information as in Aradillas-Lopez (2010), de Paula and Tang (2012), and Lewbel and Tang (2012).

Our model is closely related to index models with random coefficients. In particular, it is related to the work on the linear model in Beran, Hall and Feuerverger (1994) and Hoderlein, Klemelä and Mammen (2010). It is also related to treatment effect models as in Gautier and Hoderlein (2012), simultaneous equation models with continuous outcome as in Masten (2015), and the triangular model in Hoderlein, Holzmann and Meister (2015, HHM henceforth). Since we are considering binary dependent variables, our approach is also related to the nonparametric approach of Ichimura and Thompson (1998) as well as Gautier and Kitamura (2013). Related are also the models of Berry and Haile (2009), and Fox, Ryan, Bajari, and Kim (2012). However, nonparametric identification of the distribution of random coefficients in a simultaneous system of binary choice models has not been considered in any of these references. This paper therefore also contributes to the literature of nonparametric identification in simultaneous equation models as in Matzkin (2008), Berry and Haile (2011), Matzkin (2012), and Masten (2015).

2 The General Structural Model: Setup and Identification

In this section we introduce the basic building blocks of our model, and discuss identification of the parameter of interest. We start by providing formal notation, and clarify and discuss the assumptions. In the second subsection, we establish our main identification results and provide extensive discussions.

2.1 Setup and Assumptions

We consider a simultaneous game of complete information with two players. For instance, suppose that two firms (denoted by subscripts 1 and 2) decide whether or not to enter a market. Consider the following simultaneous equations model:

$$Y_1^* = Z_1 - (B_1 + \Delta_1 Y_2)X_1 \tag{2.1}$$

$$Y_2^* = Z_2 - (B_2 + \Delta_2 Y_1)X_2, \tag{2.2}$$

where, for $j = 1, 2$, $Y_j = 1$ if firm j enters the market and $Y_j = 0$ otherwise, X_j is a $k \times 1$ vector of covariates, possibly including a constant term, and Z_j is a scalar covariate whose coefficient is normalized to 1. B_j and Δ_j are $1 \times k$ vector random coefficients. Y_j^* is the latent utility of taking action 1, and we normalize the latent utility of taking action 0 to 0. We allow all random coefficients to depend additionally on observable covariates W, i.e., for each ω in a sample space Ω, we may write $B_j(\omega) = \tilde{B}_j(W(\omega), \omega)$, $\Delta_j(\omega) = \tilde{\Delta}_j(W(\omega), \omega)$ for some measurable maps $\tilde{B}$ and $\tilde{\Delta}$. This means that, even conditional on W, B and Δ are still random. For brevity of exposition, we will suppress the dependence on ω or the conditioning on W.

In each market, the primitives of the game determined by (B, Δ, X, Z) are assumed to be common knowledge among the players. Throughout, we assume that the observed outcome (Y_1, Y_2) is a pure strategy Nash equilibrium (PSNE) of the game. For example, when the other player's action adversely affects one's payoff, i.e. $\Delta_j X_j \geq 0, j = 1, 2$, both players entering the market is a unique pure strategy Nash equilibrium when $Z_j - (B_j + \Delta_j)X_j \geq 0, j = 1, 2$. From the literature on binary choice models, it is well known that each of the equations in (2.1)-(2.2) is only identified up to scale normalization (Ichimura and Thompson (1998)). By far the most plausible normalization is that the sign of one of the random coefficients (e.g. the coefficient on a cost shifter) is known, and hence we impose the assumption that the sign of the (possibly random) coefficient on Z_j is known to be positive and its value is normalized to 1, but see the companion paper, DHK (2014) for a more abstract approach involving a slightly more general hemisphere condition that essentially allows for a linear combination of coefficients to have a known sign. We summarize our basic assumptions on the data generating process:

Assumption 2.1. *Let $k, l \in \mathbb{N}$. For each $j = 1, 2$, let $(Z_j, X_j, W_j) \in \mathbb{R}^{1+k+l}$ be player j's observable characteristics, and let $(B_j, \Delta_j) \in \mathbb{R}^{2k}$ be random coefficients, which may depend on W. The observed outcome (Y_1, Y_2) is a pure strategy Nash equilibrium of the game characterized by (B, Δ, X, Z) with probability 1.*

For each player, the coefficient B_j captures the marginal impact of player j's own covariates X_j on the latent variable Y_j^*, typically referred to as utility (we allow for B_j to include a random intercept B_{0j}). In contrast, the strategic interaction effect Δ_j captures the impact of the other player's decision on the net utility of player j, both on the intercept and on the marginal effects, i.e., if $Y_2 = 0$, Y_1^* equals $Z_1 - B_1 X_1$, while if $Y_2 = 1$, Y_1^* equals $Z_1 - (B_1 + \Delta_1)X_1$. This interaction effect $\Delta_j X_j = \Delta_{0,j} + \Delta_{-0,j} X_{-0,j}$ comprises a strategic interaction effect on the intercept, $\Delta_{0,j} \in \mathbb{R}$, typically referred to as the interaction effect, but also an interaction effect on the responses $\Delta_{-0,j} \in \mathbb{R}^{k-1}$ to changes in observable covariates $X_{-0,j}$. In what follows, we call $\Delta = (\Delta_1, \Delta_2)$ the *strategic interaction coefficients*.

Note that Assumption 2.1 allows (B_j, Δ_j) to vary across markets. This allows us to flexibly model unobserved heterogeneity in strategic interactions across different markets. To analyze this model, we write $Y = (Y_1, Y_2)$, $Z = (Z_1, Z_2)$, $X = (X_1, X_2)$, $W = (W_1, W_2)$, $B = (B_1, B_2)$, and $\Delta = (\Delta_1, \Delta_2)$. For any random vector Q, we denote its support by S_Q. The (infinite dimensional) parameter of interest is $f_{B\Delta}$, the joint density of all random coefficients. We seek to nonparametrically identify $f_{B\Delta}$.

We will analyze this DGP in two strategic setups that will affect which conditional probabilities we will employ to identify the parameters of interest. We want to emphasize the importance of these setups for understanding the relation between the reduced form DGP and the underlying structural model of a game of complete information. Table 1 summarizes the payoffs of the game. Depending on the realizations of $\{(X_j, Z_j, W_j, B_j, \Delta_j)\}_{j=1,2}$, there exist four possible equilibrium outcomes: $(Y_1, Y_2) = (0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$. In case of multiple equilibria, one of them is selected from the set of equilibria by some selection mechanism which we do not explicitly specify. Nonetheless, nonparametric identification of the structural parameter, $f_{B\Delta}$, is possible while staying agnostic about the selection mechanism. Within each strategic setting, this can be done by investigating the conditional probabilities of the outcomes that do not involve the unknown equilibrium selection mechanism (see Tamer (2003)).

	$Y_2 = 0$ (no entry)	$Y_2 = 1$ (entry)
$Y_1 = 0$ (no entry)	$(0, 0)$	$(0, Z_2 - B_2 X_2)$
$Y_1 = 1$ (entry)	$(Z_j - B_1 X_1, 0)$	$(Z_1 - (B_1 + \Delta_1) X_1, Z_2 - (B_2 + \Delta_2) X_2)$

Table 1: The Entry Game Payoff Matrix

The leading case studied in the literature is the case where the utility of each player is adversely affected by the other players choosing action 1.[2] We require that this holds for a set of values of covariates, which we will use to identify the distribution of the random coefficients. Formally,

Assumption 2.2. *For each $w \in S_W$, there exists an open set $D(w) \subseteq \mathbb{R}^{2(k-1)}$ such that $\Delta_{0,1} + \Delta_{-0,1} x_{-0,1} \geq 0, \Delta_{0,2} + \Delta_{-0,2} x_{-0,2} \geq 0$, almost surely, for all $(x_{-0,1}, x_{-0,2}) \in D(w)$.*

As we discuss in more detail below, $D(w)$ is the set of values of non-constant covariates $(X_{-0,1}, X_{-0,2})$, which will be used in our identification argument. We require that the sign of

[2] We use the utility specification $Y_j^* = Z - (B_j + \Delta_j Y_{-j}) X_j$, and hence, the entry of the other firm adversely affects the player j's own payoff when $\Delta_j X_j \geq 0$. This specification is used for the purpose of obtaining an identification result whose relation to the existing results that use special regressors (see e.g. Lewbel, 2000) can easily be understood.

the interaction effects is known and non-negative for all $(x_{-0,1}, x_{-0,2})$ in this set. The leading example for this situation is the market entry decision of two competitive firms. We will analyze this case using an important insight of Bresnahan and Reiss (1991), who separate the outcome space into three cases, no entry $(0,0)$, duopoly $(1,1)$ and monopoly $\{(1,0),(0,1)\}$. Under Assumptions 2.1-2.2, this provides us with two separate conditional probabilities that do not involve equilibrium selections which we invert to obtain the joint distribution of $B = (B_1, B_2)$ and that of $B + \Delta = (B_1 + \Delta_1, B_2 + \Delta_2)$. From these individual pieces we recover the joint density of $\Delta = (\Delta_1, \Delta_2)$ by deconvolution when B is independent of Δ. Namely, we apply Fourier inversion to the characteristic function of the interaction effects Δ obtained from those of B and $B + \Delta$.[3]

However, we also study the case when the utility of each player is positively affected by the other players choosing action 1:

Assumption 2.3. *For each $w \in S_W$, there exists an open set $D(w) \subseteq \mathbb{R}^{2(k-1)}$ such that $\Delta_{0,1} + \Delta_{-0,1}x_{-0,1} \leq 0, \Delta_{0,2} + \Delta_{-0,2}x_{-0,2} \leq 0$, almost surely, for all $(x_{-0,1}, x_{-0,2}) \in D(w)$.*

The most illustrative example for this situation is the joint retirement of husband and wife. To analyze this case, we will again use insights of Bresnahan and Reiss (1991). In this case, they separate the outcome space into three parts: $Y = (1,0)$, $(0,1)$, and the rest. We then combine these three outcomes with insights from the deconvolution literature to obtain the joint distribution of B, as well as the two marginal distributions of the Δ_j.

Throughout, we tacitly assume that the distributions of $B = (B_1, B_2)$ and $\Delta = (\Delta_1, \Delta_2)$ are absolutely continuous with respect to some σ-finite measure and denote the (Radon-Nikodym) densities of B and Δ by f_B and f_Δ respectively. In general, this allows for continuous, discrete and mixed distributions.

We will also need enough variation in our covariates, in particular in Z. To formalize this notion, we let $S_{Q|T}$ denote the support of a random variable Q given T.

Assumption 2.4. $S_{Z|X,W} = \mathbb{R}^2$. *Moreover one of the following conditions hold. (i) The non-constant components of X are continuously distributed and have full support conditional on W, and $D(W) = \mathbb{R}^{2(k-1)}$; or (ii) The conditional support of the nonconstant components of X given W contains the open set $D(W) \subseteq \mathbb{R}^{2(k-1)}$. For $B = (B_1, B_2) = (B_{0,1}, \ldots, B_{k,1}, B_{0,2}, \ldots, B_{k,2})$, it holds that $\mathbb{E}[|B_{i,j}|^p] < \infty$ for all $i = 0, \ldots, k$, $j = 1, 2$, and $p \in \mathbb{N}$, and*

$$\lim_{p\to\infty} \frac{r^p}{p!} \mathbb{E}\Big[\Big(\sum_{i=0,\ldots,k,j=1,2} |B_{i,j}|\Big)^p \Big| W\Big] = 0, \qquad \text{for all } \ r \in \mathbb{R}, \ a.s. \tag{2.3}$$

Moreover, (2.3) *holds for $B + \Delta$ in place of B as well.*

[3] We conjecture that it is possible to incorporate Tamer's (2003) insight and use at least some of the information in the monopoly case by distinguishing between the players.

In Assumption 2.4, we require that Z has full support. This assumption is essential, as this variable plays the role of the dependent variable in the linear random coefficients model. It could be weakened to bounded support, if both the random coefficients and the covariates have bounded support, such that the support of Z_j still contains that of $B_j X_j$.

For other covariates, we consider two alternative conditions. Assumption 2.4 (i) specifies X to have full support $\mathbb{R}^{2(k-1)}$ and requires that the open set $D(W)$ (in Assumptions 2.2 and 2.3) coincides with the support. Under this condition, no additional assumption is required on the random coefficient density to achieve point identification.[4] However, full support is frequently unrealistic, and alternative assumptions have to be imposed. A case in point is Assumption 2.4 (ii), which substantially relaxes the full support benchmark assumption at the expense of additional structure on the random coefficients density. We view this as the main empirically relevant assumption. In line with Masten (2015) and Hoderlein, Holzmann, and Meister (2015), a support containing an open set suffices for point identification provided the random coefficients satisfy the moment condition in (2.3). Various commonly employed distributions satisfy this moment condition. For example, this condition is satisfied when the coefficients are normally distributed, or when they are compactly supported. Finally, note that Assumption 2.4 excludes discrete covariates or variables that show up in both players' payoffs from X. However, even in the absence of this assumption some features of the model are still (point or partially) identified, and the entire model may be point identified under additional independence assumptions, see Theorem 2.2 below.

Now we turn to the key identification condition. We assume exogeneity of covariates X, Z. Formally,

Assumption 2.5. (B, Δ) *is independent of* (X, Z) *conditional on* W.

This is the central exogeneity assumption we employ. Allowing for the independence to be conditional on additional covariates W is in line with the treatment effects literature, see, e.g., Heckman and Vytlacil (2007).

The leading special case is of course when there are no such covariates, in which case the assumption simply states that the covariates X, Z are fully independent of all unobservables in the system. This is a natural extension of assumptions made in the literature in the fixed coefficients case (e.g., Bresnahan and Reiss (1991), Tamer (2003)). Since we are explicitly considering random coefficients, our case is less restrictive than the commonly assumed full independence of a scalar additive unobservable from the covariates, because this set-up is a special case of our set-up. However, this assumption rules out correlation between (X, Z) and the random unobservables. We remark that such dependence could be admitted, if there are additional excluded instruments S, and a first stage equation is specified that allows recovery

[4] We however note that this assumption, without any further structure, has conflicts with Assumption 2.2 and 2.5 as we discuss below (Remark 2.1).

of control function residuals as in Imbens and Newey (2009). These control function residuals restore conditional independence, and serve as an important example for W.

Remark 2.1. We note that there is an interaction of Assumptions 2.2 (or 2.3), 2.4 and 2.5. This is because Assumption 2.2 (or 2.3) restricts the sign of the interaction effect $\Delta_j X_j$, Assumption 2.4 contains a support requirement for X, and Assumption 2.5 requires that Δ_j and X_j are (conditionally) independent. The consistency of these assumptions, therefore, needs to be checked in each application. In particular, when Assumption 2.4 (i) is presumed, Assumptions 2.2 (or 2.3) and 2.5 are not compatible with each other in general.[5] To see this, suppose X_j contains a single nonconstant variable with a full support. If Δ_j is nonzero, then the sign of $\Delta_j X_j$ will take positive and negative values. When Assumption 2.4 (i) is assumed, the compatibility of the assumptions therefore can only be ensured when $\Delta_{i,j} = 0, a.s.$ for $i \neq 0, j = 1, 2$. That is, the non-constant component of X_j do not enter the interaction effect (but enters the payoff through $B_j X_j$). Hence, the interaction effect is

$$\Delta_j X_j = \Delta_{0,j}. \tag{2.4}$$

In the context of market entry, this specification is called *fixed competitive effects* in Ciliberto and Tamer (2009) (CT henceforth). In this special case, Assumption 2.2 is satisfied as long as $\Delta_{0,j} \geq 0, j = 1, 2$ while Assumption 2.3 holds if $\Delta_{0,j} \leq 0, j = 1, 2$. We note, however, that this is rather a special case.

When Assumption 2.4 (i) is replaced by Assumption 2.4 (ii), consistency can be ensured through support or sign restrictions on components of Δ_j and X_j. If X_j's support is restricted, Assumption 2.4 (i) cannot be used. Therefore, we view Assumption 2.4 (ii) as the main empirically relevant assumption. We illustrate restrictions using the entry game in the airline industry as in CT, in which $-\Delta_j X_j$ (specified below) represents the opponent's impact on airline j's profits. Consider the following specification (called the *variable competitive effects* in CT):

$$\Delta_j X_j = \Delta_{0,j} + \Delta_{1,j} X_{1,j}, \tag{2.5}$$

where $X_{1,j}$ is a scalar non-negative market presence index that measures the presence of the opponent airline (e.g. # of flights operated by American Airline at an airport) in each market.[6] Suppose that the opponent's market entry itself and market presence adversely affect firm j's profit, i.e. $\Delta_{0,j} \geq 0, \Delta_{1,j} \geq 0, j = 1, 2$. These restrictions ensure Assumption 2.2 with $D(W) \subset \mathbb{R}^2_+$ and are also consistent with the independence assumption (Assumption 2.5). Since the covariate contains a non-negative variable, Assumption 2.4 (i) does not hold in this case. However, one may proceed with Assumption 2.4 (ii) by assuming that the support of the

[5] We are indebted to a referee for pointing this out.

[6] See CT for details on the definition of the market presence index. CT assume that the coefficients are non-random. Here, we allow them to be random.

nonconstant covariates contains the open set $D(W)$.

In general, Assumptions 2.2 (or 2.3) and 2.4 (ii) restrict the signs of the interaction effects only on a subset of the support of the non-constant covariates. Therefore, the support of some (or all) of these variables may contain both positive and negative values. In this case, we only use the open subset of the support on which the signs of the interaction effects are known.

Finally, we introduce an assumption that allows us to point identify the density of Δ.

Assumption 2.6. *B is independent of Δ conditional on W. The conditional characteristic function $\phi_{B|W}(\cdot|w)$ of B given $W = w$ is non-zero almost everywhere for all w.*

The condition on the characteristic function is standard in the deconvolution literature. For example, consider the commonly studied case where the strategic interaction effect enters only through the coefficient $\Delta_{0,j}$ on the intercept i.e. $\Delta_{-0,j}$ are zeros. Then, the condition is satisfied when the corresponding characteristic function of $B_{0,j}$ has isolated zeros, which holds for example when $B_{0,j}$ has bounded support, as in Carrasco and Florens (2010).

Remark 2.2. The independence assumption in Assumption 2.6 is solely used to recover f_Δ (and hence $f_{B\Delta}$). The researcher may want to consider a different assumption that allows correlation between, say, $B_{1,j}$ and $\Delta_{1,j}$ as they both capture potentially related effects of the covariate $X_{1,j}$. One alternative approach to achieve point identification would be to consider a rank invariance or rank similarity assumption as used throughout the quantile treatment effects and quantile IV literatures. This approach presumes that the direct effect B_j of a covariate and the total effect $B_j + \Delta_j$ that also incorporates the effect through strategic interactions are governed by latent variables, say U_0, U_1, which determine the ranks of these effects. The rank invariance assumption means $U_0 = U_1$, while the rank similarity only requires $U_0 \overset{d}{=} U_1$ and allows unsystematic deviations in ranks through "slippages". In the context of entry games, these variables may be interpreted as an unobserved market characteristic that determines the rank of the direct and total effects of a covariate.

If we avoid an assumption that determines the joint distribution of B_j and $B_j + \Delta_j$, this generally leads to a partial identification result in a way that is quite analogous to the distribution of treatment effects in the treatment effect literature (Heckman, Smith, and Clements (1997), Fan and Wu (2010), Gautier and Hoderlein (2012)). We provide the main intuition using the commonly studied case where $\Delta_j = \Delta_{0,j}$ under the strategic substitution assumption ($\Delta_{0,j} \geq 0$).[7] Theorem 2.1 below shows that one may recover the marginal distributions of $B_{0,j}$ and $B_{0,j} + \Delta_{0,j}$ from the conditional probabilities of the outcomes (0,0) and (1,1) respectively. Identification of the distribution of $\Delta_{0,j}$, however, requires their joint distribution. Note that identification of the distribution of $\Delta_{0,j}$ has the same structure as recovering the distribution of the treatment effect ($\Delta_{0,j}$) from the marginal distributions of the (potential)

[7] This simplification is for convenience. The bounds below apply more generally to any pair of coefficients $B_{i,j}$ and $B_{i,j} + \Delta_{i,j}$ with a sign restriction on $\Delta_{i,j}$.

outcomes ($Y_1 = B_{0,j} + \Delta_{0,j}$) and ($Y_0 = B_{0,j}$) with a monotone treatment response restriction ($Y_1 - Y_0 = \Delta_{0,j} \geq 0$). A sharp bound on the CDF of $\Delta_{0,j}$ under a sign restriction on $\Delta_{0,j}$ is derived in the recent work of Kim (2014). Hence, in this commonly studied example, one may apply his result. We leave the more general setting where Δ_j involves multiple terms for future research.

2.2 Main Identification Results

In this subsection, we provide the main identification and non-identification results. Our first main theorem establishes identification of the joint densities of all random coefficients in games of strategic substitutes and strategic complements, respectively.[8] While serving as a benchmark, this theorem, however, requires assumptions that may be problematic in a number of applications. A particularly important case is when covariates are common across players (e.g., market environments, hence called market-specific covariates). In the second part of this subsection, we establish that in this special case the joint distribution of all random coefficients is in general not point-identified. However, we also show that the joint distribution of some subcomponents of the parameter vector are point identified, as are the marginal distributions for market specific variables, and all counterfactual probabilities that only depend on the marginals. Finally, we obtain and characterize bounds on counterfactual probabilities that involve the joint distribution.

2.2.1 A General Identification Result

Theorem 2.1. *Suppose that Assumptions 2.1, 2.4-2.5 hold.*

(i) Suppose Assumption 2.2 holds. Then, $f_{B|W}$ and $f_{B+\Delta|W}$, the joint densities of B and $B+\Delta$ given W, respectively, are identified. If Assumption 2.6 holds, $f_{\Delta|W}$, the joint density of Δ given W, is also identified.

(ii) Suppose Assumption 2.3 holds. Then, $f_{(B_1,B_2+\Delta_2)}$ and $f_{(B_1+\Delta_1,B_2)|W}$, the joint densities of $(B_1, B_2+\Delta_2)$ and $(B_1+\Delta_1, B_2)$ given W, respectively, are identified. If Assumption 2.6 holds, $f_{B|W}$, the joint density of B given W, and $f_{\Delta_j|W}, j = 1, 2$, the marginal densities of Δ_1 and Δ_2 given W respectively, are also identified.

Remark 2.3. These results and Theorem 2.2 below are obtained using Proposition 1 in the appendix, which identifies the characteristic function ϕ_U of a $1 \times 2k$ random coefficient vector U from an integral equation of the following form:

$$\mathbb{P}\{Y_1 = y_1, Y_2 = y_2 | X = x, Z = z\} = \int \mathbf{1}\{z_1 < u_1 x_1\}\mathbf{1}\{z_2 < u_2 x_2\} f_U(u) du \; . \qquad (2.6)$$

[8] More precisely, in the case of strategic complements, only the joint of all B's and the marginals of the Δ_j's are identified.

In the game of strategic substitutes, it is straightforward to see from (A.1) in the appendix that $(Y_1, Y_2) = (0, 0)$ i.e., no entry, is a unique PSNE when $Z_j < B_j X_j$ for both players. If (B, Δ) is independent of (Z, X), then letting $U = B$ and $(y_1, y_2) = (0, 0)$, we obtain identification of f_B from the proposition. In the same game, $(Y_1, Y_2) = (1, 1)$ i.e. duopoly, is a unique pure strategy Nash equilibrium when $Z_j > (B_j + \Delta_j)X_j$. Hence, using the conditional probability of the duopoly outcome and applying the proposition, one may identify the distribution of $U = B + \Delta$. With f_B being identified, one may then identify f_Δ by deconvolution under Assumption 2.6.

In the game of strategic complements, $(Y_1, Y_2) = (1, 0)$ is a unique PSNE when $Z_1 > B_1 X_1$ and $Z_2 < (B_2 + \Delta_2)X_2$. Similarly, $(Y_1, Y_2) = (0, 1)$ is a unique PSNE when $Z_1 < (B_1 + \Delta_1)X_1$ and $Z_2 > B_2 X_2$. Hence, the densities of $(B_1, B_2 + \Delta_2)$ and $(B_1 + \Delta_1, B_2)$ can be identified using the conditional probabilities of these outcomes. Again by deconvolution, one may then identify the marginal densities of Δ_1 and Δ_2. However, a crucial difference from the case of strategic substitutes is that we may not identify the joint distribution of the interaction effects. This is because the conditional entry probability of the monopoly outcome $(Y_1, Y_2) = (1, 0)$ (or (0,1)) is informative about only one of the interaction effects.

Remark 2.4. Our identification result can be extended to the case with J players where $J \geq 3$. In the case of strategic substitutes with more than two players, the no entry outcome $(0, \cdots, 0)$ and "full entry" outcome $(1, \cdots, 1)$ still arise as unique equilibria. The no entry outcome for example gives the following equation:

$$\mathbb{P}\{Y_1 = 0, \cdots, Y_J = 0 | Z = z, X = x\} = \int \mathbf{1}\{z_1 < b_1 x_1\} \cdots \mathbf{1}\{z_J < b_J x_J\} f_B(b) db \ . \quad (2.7)$$

Proposition 1 can be extended in a straightforward way so that the random coefficient density can be identified from the equation above. This therefore enables us to identify the distribution of $B = (B_1, \cdots, B_J)$. With J players, however, the interaction effects become quite high-dimensional. This raises a challenge for identification. We expect that our identification strategy, which recovers f_Δ through deconvolution of $f_{B+\Delta}$ and f_B does not extend readily to this general case. However, identification of f_Δ may be possible under additional symmetry restrictions, for example, the existence of a potential function, as in Fox and Lazzati (2015).

We now extend Theorem 2.1 to allow for covariates with lower-dimensional support such as discrete variables or market specific variables. These variables do not satisfy Assumption 2.4. As we see in the next section in more detail, identifying the joint distribution of all coefficients is not generally possible in the presence of such variables. However, it is possible to identify key features of the random coefficients on those variables. To see this, suppose that each player's latent utility depends linearly on continuous covariates X_j and covariates $\check{X}_j \in \mathbb{R}^{\check{k}}$

with lower-dimensional support:

$$Y_j^* = Z_j - (B_j + \Delta_j Y_{-j})X_j - (\check{B}_j + \check{\Delta}_j Y_{-j})\check{X}_j, \tag{2.8}$$

where $\check{B}_j$ and $\check{\Delta}_j$ are $1 \times \check{k}$ random coefficients respectively. Consider the no entry outcome $(Y_1, Y_2) = (0, 0)$. Then, the utility from this outcome is $Y_j^* = Z_j - B_j X_j - \check{B}_j \check{X}_j$, $j = 1, 2$, and hence one may write $Y_j^* = Z_j - U_j X_j$, where U_j's first component (multiplying the constant term in X_j) is $B_{0,j} + \check{B}_j \check{X}_j$, and the rest of its components are the corresponding components of B_j. Once the conditional density of $B_{0,j} + \check{B}_j \check{X}_j$ (conditional on $\check{X}_j$) is identified, one may identify features of $\check{B}_j$ under additional assumptions. For example, recovering the mean of the random coefficients $\check{B}_j$ on $\check{X}_j$ from the conditional distribution of $B_{0,j} + \check{B}_j \check{X}_j$ is a linear regression problem. Alternatively, if $\check{X}_j$ is a (scalar) binary variable and $\check{B}_j$ is independent of $B_{0,j}$, one can recover the distribution of $\check{B}_j$ by another deconvolution step. Finally, we note that a market specific variable can be treated in a similar manner as covariates $\check{X}_j$ with lower-dimensional support, and imposing the condition that $\check{X}_j = \check{X}$. This last condition indicates that $\check{X}$ is a variable common to both players. Since we can recover the distribution of the variable $B_{0,j} + \check{B}_j \check{X}$ (conditional on $\check{X}$) and $B_{-0,j}$ for each player, we may apply the analysis in this section to identify the marginal densities (i.e., $f_{B_j, \check{B}_j}, f_{\Delta_j, \check{\Delta}_j}$ for each player $j = 1, 2$), as well as features that depend solely on them, e.g., certain counterfactual probabilities. Since the presence of market specific variables has important consequences for the identification of both joint and marginal densities, we postpone a more detailed discussion of this issue to the subsequent two subsections.

The following theorem summarizes identification of features of the density of $(\check{B}, \check{\Delta})$ when the latent utilities are given as in (2.8).

Theorem 2.2. *Suppose that Assumptions 2.1, 2.2, 2.4-2.6 hold with conditioning variables* $(\check{X}, W)$. *Then,*
(i) If, for any j, $\mathbb{E}[(B_{0,j}, \Delta_{0,j}, \check{B}_j, \check{\Delta}_j)|\check{X}_j, W] = \mathbb{E}[(B_{0,j}, \Delta_{0,j}, \check{B}_j, \check{\Delta}_j)|W]$, *and* $\mathbb{E}[(1, \check{X}_j')'(1, \check{X}_j')|W]$ *is positive definite almost surely, then,* $\mathbb{E}[(B_{0,j}, \Delta_{0,j}, \check{B}_j, \check{\Delta}_j)|W]$ *is identified;*
(ii) If $\check{X}_j$ *is a scalar binary random variable for* $j = 1, 2$ *such that* $\check{B} \perp \check{X}|W$, $(B_{0,1}, B_{0,2}) \perp \check{B}|W$, *and the characteristic function* $\phi_{(B_{0,1}, B_{0,2})|W}$ *is nonzero almost everywhere, the joint density* $f_{\check{B}|W}$ *is identified.*

Remark 2.5. For a semiparametric model in which $(\check{B}, \check{\Delta})$ are non-random in (2.8), Theorem 2.2 implies that these non-random coefficients are identified if $\check{X}_j$ is a random vector such that $\mathbb{E}[(1, \check{X}_j)(1, \check{X}_j)']$ is positive definite for $j = 1, 2$, or if $\check{X}_j$ is a binary random variable, for $j = 1, 2$.[9] This is because non-random coefficients automatically satisfy the (full or mean) independence assumptions. We also note that an analogous result for games of strategic complements can be obtained by similar arguments.

[9]See also Gautier and Hoderlein (2012) on binary regressors.

2.2.2 Limits to Identification: A Non-identification result for the Case of Market-Specific Covariates

To identify the joint distribution of (B_1, B_2) (or (Δ_1, Δ_2)), Assumption 2.4 requires all nonconstant components of X to be either continuously distributed and have a full support (Assumption 2.4 (i)) or to have a support containing an open set (Assumption 2.4 (ii)). This assumption is violated, for example, when X_1 and X_2 have a variable in common, e.g., a variable characterizing the market environment. As we have seen in Theorem 2.2, with additional structure, some features of the random coefficients are identifiable even in the presence of such a variable.

We show in this subsection that, in general, a violation of the support condition results in a failure of point identification of the *joint* distribution of all random coefficients. Here, we explicitly consider the settings where Assumption 2.4 is violated. For this reason, in what follows, we will treat the market-specific variable as a component of X instead of treating it as a component of another set of variables $\check{X}$ separated from X as done in the previous section.

More generally, if one of the covariates is a linear combination of the other covariates, identification of the *joint* density fails. We formally state this result below. For this, we assume that the model (2.1)-(2.2) has intercepts, say $X_{0,1} = X_{0,2} = 1$.

Theorem 2.3. *Suppose that Assumptions 2.1 and 2.5-2.6 hold. Suppose that either Assumption 2.2 or 2.3 holds. Suppose that there is a covariate $X_{i,j}$ with $1 \leq i \leq k-1$ and $j \in \{1, 2\}$ such that*

$$X_{i,j} = a'\tilde{X} \ , \tag{2.9}$$

for some $a \in \mathbb{R}^{2k-1}$ not identically equal to 0, where $\tilde{X}$ is the $2k-1$ vector of all other covariates. Then, the joint densities $f_{B|W}$ and $f_{\Delta|W}$ are not identified.

This non-identification result is due to the lack of variation generated by the exogenous covariates. Heuristically, even without Assumption 2.4, one may relate the conditional entry probabilities to the characteristic functions of the random coefficient densities. For example, in the case of strategic substitutes, the conditional probability of no entry can be linked to the following characteristic function:

$$\phi_B(t_1, t_1 x_1, t_2, t_2 x_2) = \mathbb{E}[\exp(i(B_{0,1}t_1 + B_{-0,1}t_1 x_1 + B_{0,2}t_2 + B_{-0,2}t_2 x_2))],$$
$$(t_1, t_2) \in \mathbb{R},\ (x_1, x_2) \in S_{(X_{-0,1}, X_{-0,2})}. \tag{2.10}$$

A key step toward identification of f_B is to trace out this characteristic function by varying (x_1, x_2) (together with (t_1, t_2)) on a set that is rich enough to determine ϕ_B uniquely. However, if some of the covariates have a linear relationship, one may only vary (x_1, x_2) on a low dimensional set, which is not enough for determining ϕ_B. This is indeed the case if there is a market-specific covariate shared by X_1 and X_2. Related non-identification results are established in Masten (2015) and HHM (2015), in linear simultaneous equations and triangular models, respectively (we adapt the proof in HHM (2015) to our setup). To our knowledge, this type of non-identification result in the context of discrete games is new.

Remark 2.6. Even though the *joint* distribution of all coefficients is not identified, many features of economic interest can still be identified in the presence of a market-specific covariate. As we have seen, Theorem 2.2 may be used to identify the conditional mean of the random coefficients for example. Another observation is that, provided the market-specific covariates have rich enough support, the marginal densities $f_{B_j}, f_{\Delta_j}, j = 1, 2$ can still be identified. To be precise, suppose that, for each j, all non-constant components of X_j have full support (although non-constant components of X do not have a full support jointly). This is sufficient for tracing out the characteristic function ϕ_{B_j} (or $\phi_{B_j+\Delta_j}$) for each j. One can then employ the argument used to prove Theorem 2.1 to identify the densities of these coefficients. The same result holds if we replace the full support assumption on X_j with the weaker support condition that the support of the non-constant components of X_j contains an open ball combined with an additional moment condition in (2.3), in which sums are only taken across components of each B_j for each j. We emphasize that this allows identification of counterfactual objects that depend only on the marginal densities. These objects include, for example, the conditional entry probability of player 1 given player 2's entry:

$$\mathbb{P}\{Y_1 = 1 | Z_1 = z_1^c, X_1 = x_1^c, Y_2 = 1\} = \int \mathbf{1}\{z_1^c < (b_1 + \delta_1)x_1^c\} f_{B_1}(b_1) f_{\Delta_1}(\delta_1) db_1 d\delta_1. \quad (2.11)$$

From this quantity, one may also recover derivatives, i.e., marginal counterfactual probabilities, and discrete differences in covariates.

In the next section, we provide additional sub-vector identification and partial identification results that hold even in the presence of a market specific covariate. Importantly, using these results, one can obtain bounds on counterfactuals that depend on the *joint* densities.

2.2.3 Sub-vector and Partial Identification in the Case of Market-Specific Covariates

Even though the *joint* distribution (across players) of all coefficients is not identified in the presence of market-specific covariates, many useful economic objects are still identified. In this subsection, we use the fact that even with market-specific covariates the joint distribution of some sub-vectors of random coefficients is still point identified while others are partially identified to provide partial identification results for counterfactual probabilities that involve the *joint* distribution of all coefficients.

To fix ideas, suppose that the first component of the non-constant covariates is a market specific variable and hence common across players $X_{1,1} = X_{1,2}$. Then, the joint distribution of $(B_{1,1} + B_{1,2}, \tilde{B}_1, \tilde{B}_2)$ is still identified under Assumption 2.2.[10] Here $\tilde{B}_1$ denotes B_1 without $B_{1,1}$ and $\tilde{B}_2$ denotes B_2 without $B_{1,2}$. More generally, the following corollary holds.

Corollary 2.1 (Sub-vector identification). *Suppose the assumptions of Theorem 2.1 (Assumptions 2.1 and 2.4-2.6) hold except for one covariate that is a linear combination of other covariates $X_{i,j} = a'\tilde{X}$ with $a \in \mathbb{R}^{2k-1}$, $1 \leq i \leq k-1$, and $j \in \{1, 2\}$, where a's components are $(a_{0,1}, \cdots, a_{i-1,j}, a_{i+1,j}, \cdots, a_{k-1,2})$.*

(i) If Assumption 2.2 holds, the joint density of

$$\Big(B_{0,1} + a_{0,1}B_{i,j}, \ldots, B_{i-1,j} + a_{i-1,j}B_{i,j}, B_{i+1,j} + a_{i+1,j}B_{i,j}, \ldots, B_{k-1,2} + a_{k-1,2}B_{i,j}\Big) \tag{2.12}$$

given W is identified. Furthermore, the joint density of

$$\Big(\Delta_{0,1} + a_{0,1}\Delta_{i,j}, \ldots, \Delta_{i-1,j} + a_{i-1,j}\Delta_{i,j}, \Delta_{i+1,j} + a_{i+1,j}\Delta_{i,j}, \ldots, \Delta_{k-1,2} + a_{k-1,2}\Delta_{i,j}\Big) \tag{2.13}$$

given W is identified;

(ii) If Assumption 2.3 holds, the joint density of

$$\Big(B_{0,1} + a_{0,1}B_{i,j}, \ldots, B_{i-1,j} + a_{i-1,j}B_{i,j}, B_{i+1,j} + a_{i+1,j}B_{i,j}, \ldots, B_{k-1,2} + a_{k-1,2}B_{i,j}\Big) \tag{2.14}$$

given W is identified. Furthermore, for $s = j$, the joint density of

$$\Big(\Delta_{0,s} + a_{0,s}\Delta_{i,s}, \cdots, \Delta_{i-1,s} + a_{i-1,j}\Delta_{i,s}, \Delta_{i+1,s} + a_{i+1,s}\Delta_{i,s}, \cdots, \Delta_{k-1,s} + a_{k-1,s}\Delta_{i,s}\Big) \tag{2.15}$$

given W is identified. For $s \neq j$, the joint density of

$$\Big(\Delta_{0,s} + a_{0,s}\Delta_{i,j}, \cdots, \Delta_{k-1,s} + a_{k-1,s}\Delta_{i,j}\Big) \tag{2.16}$$

given W is identified.

The market-specific covariate example discussed above corresponds to $X_{1,1} = a'\tilde{X}$ with $a_{1,2} = 1$, and all other components of a are zeros. Hence, if one's main interest is in the joint distribution of $(\tilde{B}_1, \tilde{B}_2)$, the intercepts and the coefficients on the player specific variables such as cost shifters or market presence, their joint distribution is still identified. Even further, the joint density of the strategic interaction coefficients $(\tilde{\Delta}_1, \tilde{\Delta}_2)$ on these variables is also identified.

Corollary 2.1 is also useful for obtaining bounds on counterfactual probabilities that involve joint distributions, when a market specific covariate $X_{1,1} = X_{1,2}$ is present. The key are the conditional probabilities assigned by the unidentified coefficients $(B_{1,1}, B_{1,2}, \Delta_{1,1}, \Delta_{1,2})$ to

[10]For illustrative purposes, we assume Assumption 2.2 throughout this subsection whenever we discuss this example.

two regions. For this, let us fix the identified coefficients $(\tilde{B}, \tilde{\Delta})$ to $(\tilde{b}, \tilde{\delta})$ and (X, Z) to a counterfactual value (x^c, z^c). The first region, which is denoted $R_1(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c) \subseteq \mathbb{R}^4$ under Assumption 2.2, is the region such that the model predicts (y_1, y_2) as the unique pure strategy Nash equilibrium if and only if $(B_{1,1}, B_{1,2}, \Delta_{1,1}, \Delta_{1,2})$ falls in this region. The second region, denoted $R_2(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c) \subseteq \mathbb{R}^4$, is the region such that (y_1, y_2) is predicted as one of multiple equilibria if and only if $(B_{1,1}, B_{1,2}, \Delta_{1,1}, \Delta_{1,2}) \in R_2(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c)$.[11] As in Tamer (2003) and Ciliberto and Tamer (2009), one may then derive a lower bound on the counterfactual probability through the conditional probability of the unidentified coefficients over R_1, while an upper bound can be obtained from the conditional probabilities over R_1 and R_2. In our setting, these bounds still depend on the distribution of the unidentified coefficients, which can be restricted via Corollary 2.1. It turns out that such restrictions take the form of linear constraints on the densities. The following corollary characterizes bounds that are obtained using the restrictions imposed by the model.

Corollary 2.2 (Partial Identification of Counterfactual Probabilities)**.** *Suppose the conditions of Theorem 2.1 (Assumptions 2.1 and 2.4-2.6) hold except for one covariate $X_{1,1}$, which is common between the two players so that $X_{1,1} = a\tilde{X}$ with $a_{1,2} = 1$, and all other components of a are 0. Let $w^c \in S_W$ and let (x^c, z^c) be a counterfactual value of (X, Z).*

(i) Suppose Assumption 2.2 holds. Then, $\mathbb{P}\{Y_1 = y_1, Y_2 = y_2|X = x^c, Z = z^c, W = w^c\}$ belongs to the interval $[\mathbb{P}_L(y_1, y_2|x^c, z^c, w^c), \mathbb{P}_U(y_1, y_2|x^c, z^c, w^c)]$ with

$$\mathbb{P}_t(y_1, y_2|x^c, z^c, w^c) = \int \mathbb{P}_t(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c, w^c) f_{\tilde{B}|W}(\tilde{b} \mid w^c) f_{\tilde{\Delta}|W}(\tilde{\delta} \mid w^c) d\tilde{b} d\tilde{\delta}, \ t = L, U, \tag{2.17}$$

where $\mathbb{P}_t(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c, w^c), t = L, U$ are pointwise bounds on $\mathbb{P}\{Y_1 = y_1, Y_2 = y_2|\tilde{B} = \tilde{b}, \tilde{\Delta} = \tilde{\delta}, X = x^c, Z = z^c, W = w^c\}$ that are defined as:

$$\mathbb{P}_U(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c, w^c) = \sup_{f \in \mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}} \int_{R_1(y_1,y_2|\tilde{b},\tilde{\delta},x^c,z^c) \cup R_2(y_1,y_2|\tilde{b},\tilde{\delta},x^c,z^c)} f(\theta \mid \tilde{b}, \tilde{\delta}, w^c) d\theta \tag{2.18}$$

$$\mathbb{P}_L(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c, w^c) = \inf_{f \in \mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}} \int_{R_1(y_1,y_2|\tilde{b},\tilde{\delta},x^c,z^c)} f(\theta \mid \tilde{b}, \tilde{\delta}, w^c) d\theta, \tag{2.19}$$

where $\theta = (b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2})$, and the identified set $\mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}$ for the conditional density of θ is characterized by linear equality and inequality restrictions and is given in the Appendix.

(ii) Suppose Assumption 2.3 holds. Then, $\mathbb{P}\{Y_1 = y_1, Y_2 = y_2|X = x^c, Z = z^c, W = w^c\}$ belongs to the interval $[\mathbb{P}_L(y_1, y_2|x^c, z^c, w^c), \mathbb{P}_U(y_1, y_2|x^c, z^c, w^c)]$ with

$$\mathbb{P}_t(y_1, y_2|x^c, z^c, w^c) = \int \mathbb{P}_t(y_1, y_2|\tilde{b}, x^c, z^c, w^c) f_{\tilde{B}|W}(\tilde{b} \mid w^c) d\tilde{b}, \ t = L, U, \tag{2.20}$$

[11] Bounds under Assumption 2.3 can be studied in a similar way with a slight difference due to the fact that the joint distribution of the coefficients in (2.15) and (2.16) is unidentified (only the marginals are identified).

where $\mathbb{P}_t(y_1, y_2|\tilde{b}, x^c, z^c, w^c), t = L, U$ *are pointwise bounds on* $\mathbb{P}\{Y_1 = y_1, Y_2 = y_2|\tilde{B} = \tilde{b}, X = x^c, Z = z^c, W = w^c\}$ *that are defined as:*

$$\mathbb{P}_U(y_1, y_2|\tilde{b}, x^c, z^c, w^c) = \sup_{f \in \mathcal{F}_{I,\tilde{b},w^c}} \int_{R_1(y_1,y_2|\tilde{b},x^c,z^c) \cup R_2(y_1,y_2|\tilde{b},x^c,z^c)} f(\theta \mid \tilde{b}, w^c) d\theta \tag{2.21}$$

$$\mathbb{P}_L(y_1, y_2|\tilde{b}, x^c, z^c, w^c) = \inf_{f \in \mathcal{F}_{I,\tilde{b},w^c}} \int_{R_1(y_1,y_2|\tilde{b},x^c,z^c)} f(\theta \mid \tilde{b}, w^c) d\theta, \tag{2.22}$$

where $\theta = (b_{1,1}, b_{1,2}, \delta)$*, and the identified set* $\mathcal{F}_{I,\tilde{b},w^c}$ *for the conditional density of* θ *is characterized by linear equality and inequality restrictions and is given in the Appendix.*

We note that the pointwise bounds in (2.18)-(2.19) (and (2.21)-(2.22)) are optimal values of linear programs, in which linear functionals of f are maximized subject to linear constraints imposed on the density via $\mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}$. The bounds on the counterfactual probabilities are obtained by integrating these pointwise bounds with respect to the density of $(\tilde{B}, \tilde{W})$, which is point identified by Corollary 2.1. While estimation and inference are beyond the scope of this paper, this characterization may be helpful for simplifying the computation of the bounds.

3 Conclusion and Outlook

This paper studies nonparametric identification of the joint distribution of random coefficients in binary response static games of complete information. We give conditions under which the joint distribution of random coefficients, except those on the interaction terms, is point identified. We provide stronger conditions that allow point identification of the joint density of the interaction coefficients. We also discuss various ways to extend our main identification result, and adapt it to situation which involve market specific or discrete variables.

We have focused on nonparametric identification of the density of random coefficients from uniquely predicted outcomes. An interesting direction would be to study possible efficiency gains by considering simultaneously the two integral equality restrictions obtained from the no entry and duopoly outcomes and additional integral inequality restrictions, which can be obtained from the monopoly outcomes. We pursue this in another paper that studies a setting in which the density of random coefficients are partially identified by integral equality and inequality restrictions. Another interesting direction would be to apply the developed estimation procedure to empirical examples in which heterogeneity plays an important role.

References

[1] Amemiya, T. (1974): "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables Are Truncated Normal". *Econometrica*, **42**, 999–1012.

[2] Aradillas-Lopez, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information". *Journal of Econometrics*, **157**, 409–431.

[3] Bajari, P. and H. Hong and S.P. Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information" . *Econometrica*, **78**, 1529–1568.

[4] Beran, R., A. Feuerverger, and P. Hall (1996): "On Nonparametric Estimation of Intercept and Slope in Random Coefficients Regression". *Annals of Statistics*, **2**, 2569–2592.

[5] Beresteanu, A. and I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions . *Econometrica*, **79**, 1785–1821.

[6] Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry". *Econometrica*, **60**, 889–917.

[7] Berry, S. T. and P. A. Haile (2009): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers". Working paper.

[8] Berry, S. T. and P. A. Haile (2011): "Identification in a Class of Nonparametric Simultaneous Equations Models". Working Paper.

[9] Berry, S. T. and E. Tamer (2006): "Identification in Models of Oligopoly Entry". Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume 2.

[10] Bjorn, P.A. and Q. H. Vuong (1985): "Simultaneous Equations Models for Dummy Endogenous Variables: a Game Theoretic Formulation with an Application to Labor Force Participation". Working paper.

[11] Bresnahan, T. F. and P.C. Reiss (1990): "Entry in Monopoly Market". *Review of Economic Studies*, **57**, 531–553.

[12] Bresnahan, T. F. and P.C. Reiss (1991): "Empirical Models of Discrete Games". *Journal of Econometrics*, **48**, 57–81.

[13] Carrasco, M. and J.P. Florens (2010): "A Spectral Method for Deconvolving a Density". *Econometric Theory*, **27**, 546–581.

[14] Chesher, A. and A. M. Rosen (2012): "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification". CEMMAP Working Paper

[15] Ciliberto, F. and E. Tamer (2009): "Market structure and multiple equilibria in airline markets". *Econometrica*, **77**, 1791–1828.

[16] Cuesta-Albertos, Juan Antonio and Fraiman, Ricardo and Ransford, Thomas (2007): "A sharp form of the Cramér-Wold theorem. *Journal of Theoretical Probability*, **20**, 201–209.

[17] de Paula, A., and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games With Incomplete Information". *Econometrica*, **80**, 143–172.

[18] Dunker, F., S. Hoderlein, and H. Kaido (2014): "Random Coefficients in Static Games of Complete Information". Working Paper

[19] Fan, Y. and J. Wu (2010): “Partial Identification of the Distribution of Treatment Effects in Switching Regime Models and its Confidence Sets”. *Review of Economic Studies* **77**, 1002–1041.

[20] Fox, J.T. and N. Lazzati (2015): “Identification of Discrete Games and Choice Models for Bundles”. Working paper.

[21] Fox, J.T., S.P. Ryan, P. Bajari, and K. Kim (2012): “The random coefficients logit model is identified”. *Journal of Econometrics*, **166**, 204–212.

[22] Gautier, E., and S. Hoderlein (2012): “Estimating the Distribution of Treatment Effects”. CeMMAP Working Paper.

[23] Gautier, E., and Y. Kitamura (2013): “Nonparametric Estimation in Random Coefficients Binary Choice Models”. *Econometrica*, **81**, 581–607.

[24] Heckman, J. J. (1978): “Dummy Endogenous Variables in a Simultaneous Equation System”. *Econometrica*, **46**, 931–959.

[25] Heckman, J. J., J. Smith, and N. Clements (1997): “Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts”. *Review of Economic Studies*, **64**, 487–535.

[26] Heckman, J. J., and E. J. Vytlacil (2007): “Econometric evaluation of social programs, part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments. Handbook of Econometrics, Vol.6B, Chapter 71.

[27] Hoderlein, S., J. Klemelä, and E. Mammen (2010): “Analyzing the Random Coefficient Model Nonparametrically”. *Econometric Theory*, **26**, 804–837.

[28] Hoderlein, S., H. Holzmann, A. Meister (2015): “The Triangular Model with Random Coefficients”. Working Paper

[29] Ichimura, H., and T. Thompson (1998): “Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution”. *Journal of Econometrics*, **86**, 269–295.

[30] Imbens, G. and W. Newey (2009): “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity”. *Econometrica*, **77**, 1481–1512.

[31] Kim, J. H. (2014): “Identifying the Distribution of Treatment Effects under Support Restrictions ”. Working Paper

[32] Kline, B. (2015): “Identification of Complete Information Games”. *Journal of Econometrics*, **189**, 117–131.

[33] Kline, B. and Tamer, E. (2012): “Bounds for Best Response Functions in Binary Games”. *Journal of Econometrics*, **166**, 92–105.

[34] Lewbel, A. (2000): “Semiparametric Qualitative Response Model Estimation with Unknown heteroskedasticity or Instrumental Variables”. *Journal of Econometrics*, **97**, 145–177.

[35] Lewbel, A., and X. Tang (2012): "Identification and Estimation of Games with Incomplete Information using Excluded Regressors". Working Paper

[36] Masten, M. (2015): "Random Coefficients on Endogenous Variables in Simultaneous Equations Models". Working Paper

[37] Matzkin, R. L. (2008): "Identification in Nonparametric Simultaneous Equations Models". *Econometrica*, **76**, 945â€"978.

[38] Matzkin, R. L. (2012): "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity". *Journal of Econometrics*, **166**, 106â€"15.

[39] Rudin, W. (1974): Real and Complex Analysis, Wiley.

[40] Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria". *Review of Economic Studies*, **70**, 147–165.

A Appendix

Recall that $B = (B_1, B_2)$ and $\Delta = (\Delta_1, \Delta_2)$ are $1 \times 2k$ random coefficient vectors of interest. Our principal objective is to identify, under different game specifications, various densities such as $f_{B|W}$ and $f_{\Delta|W}$, where W is either a null or nonnull vector of regressors. Proposition 1 below will be the main result we use to achieve this objective. In this appendix, after establishing some preliminary lemmas, we prove Proposition 1, which we then use to prove the main theorems in the paper, Theorem 2.1 and Theorem 2.2.

Let $U = (U_1, U_2)$ denote a general $1 \times 2k$ random coefficient vector. For example, U may stand for B, $B + \Delta$, or similar objects. Let (y_1, y_2) denote possible realizations of the binary outcomes (Y_1, Y_2). Note that the conditional entry probability $\mathbb{P}\{Y_1 = y_1, Y_2 = y_2 | X = x, Z = z, W = w\}$ is identified from the data. Moreover, this quantity is equal to

$$\begin{aligned} &\int \mathbf{1}\{z_1 < u_1 x_1\}\mathbf{1}\{z_2 < u_2 x_2\} f_{U|XZW}(u|x,z,w)du \quad (y_1, y_2) = (0,0),\ U = B \\ &\int \mathbf{1}\{z_1 > u_1 x_1\}\mathbf{1}\{z_2 > u_2 x_2\} f_{U|XZW}(u|x,z,w)du \quad (y_1, y_2) = (1,1),\ U = B + \Delta \end{aligned} \tag{A.1}$$

when Assumption 2.2 (strategic substitutes) holds. This is because $(Y_1, Y_2) = (0, 0)$ is a unique pure strategy Nash equilibrium when $Z_1 - B_1 X_1 < 0$ and $Z_2 - B_2 X_2 < 0$, and $(Y_1, Y_2) = (1, 1)$ is a unique pure strategy Nash equilibrium when $Z_1 - B_1 X_1 > 0$ and $Z_2 - B_2 X_2 > 0$. Similarly, the conditional entry probability is equal to

$$\begin{aligned}
&\int \mathbf{1}\{z_1 < u_1x_1\}\mathbf{1}\{z_2 > u_2x_2\}f_{U|XZW}(u|x,z,w)du \quad (y_1,y_2)=(0,1),\ U=(B_1,B_2+\Delta_2) \\
&\int \mathbf{1}\{z_1 > u_1x_1\}\mathbf{1}\{z_2 < u_2x_2\}f_{U|XZW}(u|x,z,w)du \quad (y_1,y_2)=(1,0),\ U=(B_1+\Delta_1,B_2),
\end{aligned} \tag{A.2}$$

when Assumption 2.3 (strategic complements) holds.

Proposition 1. *Suppose $S_{Z|X,W}=\mathbb{R}^2$ and $U \perp (X,Z)|W$. Suppose $\mathbb{P}\{Y_1=y_1,Y_2=y_2|X=x,Z=z,W=w\}$ is equal to the quantities in* (A.1) *or* (A.2). *Suppose one of the following conditions hold. (i) The nonconstant components of X have support $\mathbb{R}$; or (ii) The support of X contains an open set D. The moments of all components are finite $\mathbb{E}[|U_{i,j}|^p]<\infty$ for all $i=1,\dots,k, j=1,2$ and $p\in\mathbb{N}$. In addition, for any $r>0$,*

$$\lim_{p\to\infty}\frac{r^p}{p!}\mathbb{E}\Big[\Big(\sum_{i=0,\cdots,k,j=1,2}|U_{i,j}|\Big)^p\Big|W\Big]=0,\ a.s.$$

Then $\phi_{U|W}$, the characteristic function of U given W, is identified.

Before we prove Proposition 1, we establish several useful lemmas. For ease of exposition, we prove all results for the special case $(y_1,y_2)=(0,0)$ and W null. When W is nonnull, all results generalize immediately by adding the conditioning variable in obvious places. The case $(y_1,y_2)=(0,0)$ corresponds to the case $U=B$. The proof for this case is easily adapted to cover the other cases, as we show.

Lemma A.1 below shows that

$$\begin{aligned}
&\mathbb{P}\{Y_1=0,Y_2=0 \mid Z=z,X=x\} \\
&\qquad =\int_{z_2}^{\infty}\int_{z_1}^{\infty}\left[\int \mathbf{1}\{s_1=u_1x_1\}\mathbf{1}\{s_2=u_2x_2\}f_U(u)d\lambda_1 d\lambda_2\right]ds_1ds_2\,.
\end{aligned}\tag{A.3}$$

where $d\lambda_t$ denotes an increment along the hyperplane $s_t=u_tx_t$, $t=1,2$.

It is convenient to introduce some compact notation. Write $\{Y=0\}$ for the event $\{Y_1=0,Y_2=0\}$ and $\int_z^\infty$ for $\int_{z_2}^\infty\int_{z_1}^\infty$. Let $S=(S_1,S_2)$. Note that when $Y_1=Y_2=0$, $S_j=B_jX_j$, $j=1,2$. Write $\{S=BX\}$ for the event $\{S_1=B_1X_1,S_2=B_2X_2\}$. Next, write $\{Z<S\}$ for $\{Z_1<S_1,Z_2<S_2\}$ and $\{Z<BX\}$ for $\{Z_1<B_1X_1,Z_2<B_2X_2\}$. Note that $\{Z<BX\}=\{Z<S\}\{S=BX\}$. Finally, write $d\lambda$ for $d\lambda_1 d\lambda_2$.

Lemma A.1. *Suppose the conditions of Proposition 1 hold. Then*

$$\mathbb{P}\{Y=0 \mid Z=z,X=x\}=\int_z^\infty\left[\int\{s=ux\}f_B(u)d\lambda\right]ds\,. \tag{A.4}$$

PROOF. By (A.1),

$$\mathbb{P}\{Y = 0 \mid Z = z, X = x\} = \int_{B|ZX} \{z < bx\} f_{B|ZX}(b \mid z, x) db\,.$$

By the law of iterated expectations and the fact that $\{z < bx\} = \{z < s\}\{s = bx\}$,

$$\begin{aligned}
\mathbb{P}(Y = 0 \mid Z = z, X = x) &= \mathbb{E}_{S|ZX} \mathbb{P}(Y = 0 \mid S = s, Z = z, X = x) \\
&= \int_{S_{S|ZX}} \{z < s\} \left[\int \{s = bx\} f_{B|SZX}(b \mid s, z, x) d\lambda \right] f_{S|ZX}(s \mid z, x) ds \\
&= \int_z^\infty \left[\int \{s = bx\} f_{B|SZX}(b \mid s, z, x) d\lambda \right] f_{S|ZX}(s \mid z, x) ds\,.
\end{aligned}$$

Consider $f_{B|SZX}(b \mid s, z, x) f_{S|ZX}(s \mid z, x)$. Drop the subscripts. We get that

$$\begin{aligned}
f(b \mid s, z, x) f(s \mid z, x) &= \frac{f(b, s, z, x)}{f(s, z, x)} \frac{f(s, z, x)}{f(z, x)} \\
&= \frac{f(b, s \mid z, x) f(z, x)}{f(z, x)} \\
&= f(b, s \mid z, x)\,.
\end{aligned}$$

By hypothesis, $B \perp (Z, X)$. It follows that $(B, S) \perp Z \mid X$ and so $f(b, s \mid z, x) = f(b, s \mid x)$. Finally, note that

$$\begin{aligned}
f(b, s \mid x) &= f(s \mid b, x) f(b \mid x) \\
&= \mathbf{1}\{s = bx\} f(b)\,.
\end{aligned}$$

The last equality follows from the fact $S = BX$ and so $f(s \mid b, x)$ is a single point mass distribution; $f(b \mid x) = f(b)$ since $B \perp X$. Statement (A.4) follows. ☐

Let ∂_z denote $\partial_{z_1}\partial_{z_2}$. Define $\psi_{00}(z, x) = \partial_z \mathbb{P}(Y = y \mid Z = z, X = x)$ and let $\{z = ux\}$ denote $\{z_1 = u_1 x_1, z_2 = u_2 x_2\}$.

Lemma A.2. *Suppose the conditions of Proposition 1 hold. Then*

$$\psi_{00}(z, x) = \int \{z = bx\} f_B(b) d\lambda\,.$$

PROOF. If, for each $c \in \mathbb{R}$, $\int_c^\infty g(t)dt = \int_c^\infty h(t)dt$, then $g(t) = h(t)$ for almost all $t \in \mathbb{R}$. This simple fact is easily proved and is used repeatedly in the proof.

Note that (A.1) says that $\mathbb{P}\{Y_1 = 0, Y_2 = 0 \mid X = x, Z = z\}$ equals

$$\int\int \mathbf{1}\{z_1 < b_1 x_1\} \mathbf{1}\{z_2 < b_2 x_2\} f_{B_1 B_2|XZ}(b_1, b_2 \mid x, z) db_1 db_2\,.$$

Since the two indicator functions depend on (B_1, B_2) only through (S_1, S_2) where $S_i = B_i x_i$, $i = 1, 2$, we get that $\mathbb{P}\{Y_1 = 0, Y_2 = 0 \mid X = x, Z = z\}$ is also equal to

$$\begin{aligned} & \int\int \mathbf{1}\{z_1 < s_1\}\mathbf{1}\{z_2 < s_2\} f_{S_1S_2|XZ}(s_1, s_2 \mid x, z) ds_1 ds_2 \\ = & \int_{z_2}^{\infty} \left[\int_{z_1}^{\infty} f_{S_1S_2|XZ}(s_1, s_2 \mid x, z) ds_1 \right] ds_2 \qquad \text{(A.5)} \\ = & \int_{z_2}^{\infty} g(s_2) ds_2 \,. \qquad \text{(A.6)} \end{aligned}$$

Lemma A.1 shows that $\mathbb{P}\{Y_1 = 0, Y_2 = 0 \mid X = x, Z = z\}$ is also equal to

$$\begin{aligned} & \int_{z_2}^{\infty} \left[\int_{z_1}^{\infty} \left[\int \mathbf{1}\{s_1 = b_1 x_1\}\mathbf{1}\{s_2 = b_2 x_2\} f_{B_1B_2}(b_1, b_2) d\lambda_1 d\lambda_2 \right] ds_1 \right] ds_2 \\ = & \int_{z_2}^{\infty} h(s_2) ds_2 \,. \qquad \text{(A.7)} \end{aligned}$$

Since $f_{S_1S_2|XZ}(s_1, s_2 \mid x, z)$ is a density, it follows from (A.5) that $g(s_2)$ is L_1-integrable. By the simple fact stated at the beginning of the proof, (A.6), and (A.7), $g(s_2) = h(s_2)$ for almost all s_2. It follows that $h(s_2)$ is L_1-integrable. Deduce from the simple fact, (A.6), (A.7), and the Lebesgue Differentiation Theorem (Rudin (1974),Theorem 8.17) that

$$\begin{aligned} \partial_{z_2} \left[\int_{z_2}^{\infty} g(s_2) ds_2 \right] &= -g(z_2) \\ &= -\int_{z_1}^{\infty} f_{S_1S_2|XZ}(s_1, z_2 \mid x, z) ds_1 \qquad \text{(A.8)} \\ &= -\int_{z_1}^{\infty} \left[\int \mathbf{1}\{s_1 = b_1 x_1\}\mathbf{1}\{z_2 = b_2 x_2\} f_{B_1B_2}(b_1, b_2) d\lambda_1 d\lambda_2 \right] ds_1 \text{(A.9)} \\ &= -h(z_2) \\ &= \partial_{z_2} \left[\int_{z_2}^{\infty} h(s_2) ds_2 \right] . \end{aligned}$$

Now repeat the argument for the partial with respect to z_1. Define

$$\begin{aligned} \tilde{g}(s_1) &= -f_{S_1S_2|XZ}(s_1, z_2 \mid x, z) \\ \tilde{h}(s_1) &= -\int \mathbf{1}\{s_1 = b_1 x_1\}\mathbf{1}\{z_2 = b_2 x_2\} f_{B_1B_2}(b_1, b_2) d\lambda_1 d\lambda_2 \,. \end{aligned}$$

Since $f_{S_1S_2|XZ}(s_1, s_2 \mid x, z)$ is a density, it follows from (A.8) that $\tilde{g}(s_1)$ is L_1-integrable. By the simple fact, (A.8), and (A.9), we get that $\tilde{g}(s_1) = \tilde{h}(s_1)$ for almost all s_1. It follows that $\tilde{h}(s_1)$ is L_1-integrable. Deduce from the simple fact, (A.8), (A.9), and the Lebesgue Differentiation

Theorem that

$$\begin{aligned}
\partial_{z_1}\left[\int_{z_1}^{\infty}\tilde{g}(s_1)ds_1\right] &= -\tilde{g}(z_1) \\
&= f_{S_1S_2|XZ}(z_1,z_2 \mid x,z) \\
&= \int \mathbf{1}\{z_1=b_1x_1\}\mathbf{1}\{z_2=b_2x_2\}f_{B_1B_2}(b_1,b_2)d\lambda_1 d\lambda_2 \\
&= -\tilde{h}(z_1) \\
&= \partial_{z_1}\left[\int_{z_1}^{\infty}\tilde{h}(s_1)ds_1\right] \\
&= \partial_{z_1}\partial_{z_2}\mathbb{P}\{Y_1=0,Y_2=0 \mid X=x,Z=z\} = \psi_{00}(z,x)
\end{aligned}$$

which is what we wanted to show. □

Define $\|x\| = \|x_1\|\|x_2\|$.

Lemma A.3. *Suppose the conditions of Proposition 1 hold. Then*

$$\int_{S_{Z|X}} \exp(itz)\psi_{00}(z,x)dz = \|x\| \int_{\mathbb{R}^k\times\mathbb{R}^k} \exp(itbx)f_B(b)db\,.$$

Proof. By Lemma A.2,

$$\int_{S_{Z|X}} \exp(itz)\psi_{00}(z,x)dz = \int_{S_{Z|X}}\int \exp(itz)\{z=bx\}f_B(b)d\lambda dz\,.$$

To evaluate the integral on the right, we start by parametrizing the hyperplanes $z_1 = b_1x_1$ and $z_2 = b_2x_2$. Let $B^{(1)} = \{B_1^{(1)},\ldots,B_{k-1}^{(1)}\}$ denote an orthonormal basis for the subspace $0 = b_1x_1$. Likewise, let $B^{(2)} = \{B_1^{(2)},\ldots,B_{k-1}^{(2)}\}$ denote an orthonormal basis for the subspace $0 = b_2x_2$. Note that each $B_i^{(j)}$ is orthogonal to x_j. Also, in keeping with previous notation, $B_i^{(j)}$ is a $1\times k$ vector.

For $j = 1,2$, let $M_{B^{(j)}}$ denote the $(k-1)\times k$ matrix formed by stacking the $k-1$ basis vectors $B_1^{(j)},\ldots,B_{k-1}^{(j)}$ on top of each other. Hence, for every row vector $r_j \in \mathbb{R}^{k-1}$ we have $b_j \equiv z_jx_j'\|x_j\|^{-2} + r_jM_{B^{(j)}}$ on the hyperplane $z_j = b_jx_j$. We reformulate the integral as follows:

$$\begin{aligned}
&\int_{S_{Z|X}}\int \exp(itz)\{z=bx\}f_B(b)d\lambda dz \\
&= \int_{S_{Z|X}}\int_{\mathbb{R}^{k-1}\times\mathbb{R}^{k-1}} \exp(itz)f_B(z_1x_1'\|x_1\|^{-2}+r_1M_{B^{(1)}}, z_2x_2'\|x_2\|^{-2}+r_2M_{B^{(2)}})dr_1dr_2dz \\
&= \int_{S_{Z_2|X_2}\times\mathbb{R}^{k-1}}\int_{S_{Z_1|X_1}\times\mathbb{R}^{k-1}} \exp(itz)f_B(z_1x_1'\|x_1\|^{-2}+r_1M_{B^{(1)}}, z_2x_2'\|x_2\|^{-2}+r_2M_{B^{(2)}})d(r_1,z_1,r_2,z_2)\,.
\end{aligned}$$

The notation $d(r_1,z_1,r_2,z_2)$ in the last line emphasizes that, by the Tonelli-Fubini Theorem, the order of integration does not matter. We take $d(r_1,z_1,r_2,z_2) = dr_1dz_1dr_2dz_2$.

For $j = 1, 2$, construct the $k \times k$ matrix M_j consisting of the $1 \times k$ vector $x_j' \|x_j\|^{-2}$ stacked on top of the $(k-1) \times k$ matrix $M_{B^{(j)}}$. Consider the transformation $b_j = (z_j, r_j) M_j$. Note that M_j is equal to an orthonormal matrix with its first row multiplied by $\|x_j\|^{-1}$. It follows from elementary properties of determinants that the Jacobian of the transformation equals $\det(M_j^{-1}) = \pm \|x_j\|$. Deduce that $dr_j dz_j = \|x_j\| db_j$.

Note that $z = bx$ is on the hyperplanes. For compactness, formally write dr for $dr_1 dr_2$, db for $db_1 db_2$, and $zx'\|x\|^{-2} + rM_B$ for $z_1 x_1' \|x_1\|^{-2} + r_1 M_{B^{(1)}}, z_2 x_2' \|x_2\|^{-2} + r_2 M_{B^{(2)}}$. Put it all together, applying $S_{Z_j|X_j} = \mathbb{R}$, to get

$$\begin{aligned}
\int_{S_{Z|X}} \exp(itz) \psi_{00}(z, x) dz &= \int_{S_{Z|X}} \int \exp(itz) \{z = bx\} f_B(b) d\lambda dz \\
&= \int_{S_{Z|X}} \int_{\mathbb{R}^{k-1} \times \mathbb{R}^{k-1}} \exp(itz) f_B(zx'\|x\|^{-2} + rM_B) dr dz \\
&= \int_{S_{Z_2|X_2} \times \mathbb{R}^{k-1}} \int_{S_{Z_1|X_1} \times \mathbb{R}^{k-1}} \exp(itz) f_B(zx'\|x\|^{-2} + rM_B) \|x\| dr_1 dz_1 dr_2 dz_2 \\
&= \|x\| \int_{\mathbb{R}^k \times \mathbb{R}^k} \exp(itbx) f_B(b) db\,.
\end{aligned}$$

□

Lemma A.4. *Let $B = (B_1, \ldots, B_k)$ be a k-dimensional random vector. Assume that the moments of all components are finite, that is, $\mathbb{E}[|B_i|^p] < \infty$ for all $i = 1, \ldots, k$ and $p \in \mathbb{N}$. In addition, assume that for any $r > 0$,*

$$0 = \lim_{p \to \infty} \frac{r^p}{p!} \mathbb{E}[(|B_1| + |B_2| + \ldots + |B_k|)^p].$$

Finally, assume that the characteristic function ϕ_B of B is known on some open neighborhood $\mathcal{U} \subset \mathbb{R}^k$. Then ϕ_B is identified on $\mathbb{R}^k$.

Proof. We follow a similar proof strategy as in HHM Lemma A.2. The characteristic function ϕ_B can be approximated by the p-th Taylor polynomial for some point $b_0 \in \mathcal{U}$. The Taylor remainder for some point $b \in \mathbb{R}^k$ is bounded by

$$(R_p \phi_B)(b; b_0) \leq \sum_{\alpha \in \mathbb{N}^k, |\alpha| = p+1} \frac{(b - b_0)^\alpha}{\alpha!} \|D^\alpha \phi_B\|_\infty .$$

In this formula the multi-index notation is used with respect to α. This means $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \in \mathbb{N}^k$, $|\alpha| := \sum_{i=1}^k \alpha_i$, $\alpha! := \prod_{i=1}^k \alpha_i!$, and

$$D^\alpha \phi_B = \frac{\partial^{|\alpha|} \phi_B}{\partial b_1^{\alpha_1} \partial b_2^{\alpha_2} \ldots \partial b_k^{\alpha_k}}.$$

Using,

$$\|D^{\alpha}\phi_B\|_{\infty} \leq \mathbb{E}[|B^{\alpha}|] = \mathbb{E}[|B_1^{\alpha_1}B_2^{\alpha_2}\ldots B_k^{\alpha_k}|] \leq \mathbb{E}[|B_1|^{\alpha_1}|B_2|^{\alpha_2}\ldots|B_k|^{\alpha_k}]$$

we get

$$\begin{aligned}
(R_p f)(b;b_0) &\leq \|b-b_0\|_{\infty}^{p}\,\mathbb{E}\left[\sum_{\alpha\in\mathbb{N}^k,|\alpha|=p+1}\frac{|B_1|^{\alpha_1}|B_2|^{\alpha_2}\ldots|B_k|^{\alpha_k}}{\alpha!}\right] \\
&\leq \|b-b_0\|_{\infty}^{p}\,\mathbb{E}\left[(p!)^{-1}\left(|B_1|+|B_2|+\ldots+|B_k|\right)^{p}\right] \\
&\leq \frac{\|b-b_0\|_{\infty}^{p}}{p!}\,\mathbb{E}\left[\left(|B_1|+|B_2|+\ldots+|B_k|\right)^{p}\right].
\end{aligned}$$

Hence, the Taylor approximation converges point-wise to ϕ_B on $\mathbb{R}^k$. Thereby ϕ_B is identified. □

We are now prepared to prove Proposition 1.

PROOF OF PROPOSITION 1: Fix $t = (t_1, t_2) \in \mathbb{R}^2$ and consider the following conditional expectation:

$$\begin{aligned}
\phi(t \mid x) &\equiv \mathbb{E}_{Z|X=x}\left[\frac{\exp(itZ)\psi_{00}(Z,x)}{f_{Z|X}(Z \mid x)}\right] \\
&= \int_{S_{Z|X}}\left[\frac{\exp(itz)\psi_{00}(z,x)}{f_{Z|X}(z \mid x)}\right] f_{Z|X}(z \mid x)dz \\
&= \int_{S_{Z|X}} \exp(itz)\psi_{00}(z,x)dz \\
&= \|x\| \int_{\mathbb{R}^k\times\mathbb{R}^k} \exp(itbx) f_B(b)db\,.
\end{aligned}$$

The last equality follows from Lemma A.3.

Note that the integral in the last expression can be written as $\mathbb{E}\left[\exp(i(B_1\sigma_1 + B_2\sigma_2))\right]$ where $\sigma_1 = t_1x_1$ and $\sigma_2 = t_2x_2$. This expectation equals $\phi_B(\sigma)$, the characteristic function of B evaluated at $\sigma = (\sigma_1, \sigma_2)$. Since we can identify both $\psi_{00}(z,x)$ and $f_{Z|X}(z \mid x)$ for each (z,x), we can identify $\phi(t \mid x)$ for each t and x. When (i) holds, the map from x to σ is one-to-one and onto $S_\sigma = \mathbb{R}^k \times \mathbb{R}^k$. Deduce that we can identify $\phi_B(\sigma)$ for each σ. Alternatively, when (ii) holds, consider the set $D_\sigma = \{\sigma : \sigma_j = t_jx_j, j = 1,2, t \in T, x \in D\}$, where $T \subset \mathbb{R}^2$ is an open set. For our purposes, it suffices to take $T = (\epsilon, \infty)^2$ for some $\epsilon > 0$. We may identify $\phi_B(\sigma)$ for each $\sigma \in D_\sigma$. Note that, for each j, t_j equals the first component $\sigma_{1,j}$ of σ_j, and $x_j = \sigma_j/\sigma_{1,j}$. Therefore, the inverse map $\sigma \mapsto (t,x)$ is well-defined and continuous. By (ii), $T \times D$ being open, and $\sigma \mapsto (t,x)$ being continuous, D_σ is open. By (ii) and Lemma A.4, ϕ_B is uniquely determined by its restriction to an open set. Deduce that ϕ_B is identified for each σ. □

We are now ready to prove the main identification results. Again, we prove all results for the case W null. We therefore denote the open set in Assumptions 2.2 (or 2.3) and 2.4 by D. When W is nonnull, all results generalize immediately by adding the conditioning variable in obvious places.

PROOF OF THEOREM 2.1: We start by proving part (i) of the theorem. Let $D = \mathbb{R}^{2(k-1)}$ if Assumption 2.4 (i) holds. Let D be an open subset of $S_{(X_{-0,1},X_{-0,2})}$ if Assumption 2.4 (ii) holds. For each $j = 1, 2$, let $x_j = (1, x_{-0,j})$, where $(x_{-0,1}, x_{-0,2}) \in D$. Consider the case $(Y_1, Y_2) = (0, 0)$. By Assumptions 2.1-2.2, it is straightforward to check that $(Y_1, Y_2) = (0, 0)$ is a unique PSNE when $Z_1 < B_1x_1$ and $Z_2 < B_2x_2$. Hence, equation (A.1) holds with $(y_1, y_2) = (0, 0)$ and $U = B$. Assumptions 2.4-2.5 then ensure the remaining conditions of Proposition 1. By Proposition 1, ϕ_B, the characteristic function of B is identified. It follows that we can identify $f_B(b)$ by Fourier inversion:

$$f_B(b) = \int_{S_\sigma} \exp(-i(b_1\sigma_1 + b_2\sigma_2))\phi_B(\sigma)d\sigma\,. \tag{A.10}$$

Hence, $f_B(b)$ is identified.

Next, we consider the case $(Y_1, Y_2) = (1, 1)$. By Assumptions 2.1-2.2, $(Y_1, Y_2) = (1, 1)$ is a unique PSNE when $Z_1 > (B_1+\Delta_1)x_1$ and $Z_2 > (B_2+\Delta_2)x_2$. Let $\{Y = 1\}$ denote $\{Y_1 = 1, Y_2 = 1\}$ and let $\int_{-\infty}^{z}$ denote $\int_{-\infty}^{z_2}\int_{-\infty}^{z_1}$. Let $U = B + \Delta$ and $u = (u_1, u_2) = (b_1 + \delta_1, b_2 + \delta_2)$. Let $\{s = ux\}$ denote $\{s_1 = u_1x_1, s_2 = u_2x_2\}$. Mimic the argument in the proof of Lemma A.1, making the obvious changes to get $\mathbb{P}\{Y = 1 \mid X = x, Z = z\} = \int_{\infty}^{z}\left[\int\{s = ux\}f_U(u)\right]ds$. Define $\psi_{11}(x, z) = \partial_z\mathbb{P}(Y = 1 \mid X = x, Z = z)$ and let $\{z = ux\}$ denote $\{z_1 = u_1x_1, z_2 = u_2x_2\}$. Mimic the argument in the proof of Lemma A.2 to get $\psi_{11}(x, z) = \int\{z = ux\}f_U(u)d\lambda$. Mimic the argument in the proof of Lemma A.3 to get $\int_{S_{Z|X}} \exp(itz)\psi_{11}(x, z)dz = \|x\|\int_{\mathbb{R}^k\times\mathbb{R}^k} \exp(itux)f_U(u)du$. Finally, mimic the argument in the proof of Proposition 1 to get that $\phi_U = \phi_{B+\Delta}$ is identified.

If Assumption 2.6 holds, $\phi_U(\sigma) = \phi_B(\sigma)\phi_\Delta(\sigma)$ and $\phi_\Delta(\sigma)$ is well-defined almost everywhere. Deduce from this and ϕ_B being identified by the previous step that $\phi_\Delta(\sigma)$ is identified. By Fourier inversion, $f_\Delta(\delta)$ is identified. This proves part (i).

Next, we prove part (ii) of the theorem. Consider the case $(Y_1, Y_2) = (1, 0)$. By Assumptions 2.1 and 2.3, $(Y_1, Y_2) = (1, 0)$ is a unique PSNE when $Z_1 > B_1x_1$ and $Z_2 < (B_2 + \Delta_2)x_2$. Let $\{Y = (1, 0)\}$ denote $\{Y_1 = 1, Y_2 = 0\}$. Let $U = (B_1, B_2 + \Delta_2)$ and $u = (u_1, u_2) = (b_1, b_2 + \delta_2)$. Let $\{s = ux\}$ denote $\{s_1 = u_1x_1, s_2 = u_2x_2\}$. Mimic the argument in the proof of Lemma A.1, making the obvious changes to get $\mathbb{P}\{Y = (1, 0) \mid X = x, Z = z\} = \int_{z_2}^{\infty}\int_{-\infty}^{z_1}\left[\int\{s = ux\}f_U(u)\right]ds$. Define $\psi_{10}(x, z) = \partial_z\mathbb{P}(Y = (1, 0) \mid X = x, Z = z)$ and let $\{z = ux\}$ denote $\{z_1 = u_1x_1, z_2 = u_2x_2\}$. Mimic the argument in the proof of Lemma A.2 to get $\psi_{10}(x, z) = -\int\{z = ux\}f_U(u)d\lambda$. Mimic the argument in the proof of Lemma A.3 to get $\int_{S_{Z|X}} \exp(itz)\psi_{01}(x, z)dz = -\|x\|\int_{\mathbb{R}^k\times\mathbb{R}^k} \exp(itux)f_U(u)du$. Finally, mimic the argument in the proof of Proposition 1 to get that $\phi_U(\sigma)$ is identified.

Now, let $V = (B_1 + \Delta_1, B_2)$. If we consider the case $(Y_1, Y_2) = (0, 1)$ and mimic the argument in the last paragraph, we get that $\phi_V(\sigma)$, the characteristic function of V, is identified. Therefore, the characteristic functions of $B_2 + \Delta_2$ and B_2 are identified as follows:

$$\phi_{B_2+\Delta_2}(\sigma_2) = \phi_U(0, \sigma_2), \quad \text{and} \quad \phi_{B_2}(\sigma_2) = \phi_V(0, \sigma_2). \tag{A.11}$$

By Assumption 2.6, $\phi_{B_2+\Delta_2}(\sigma_2) = \phi_{B_2}(\sigma_2)\phi_{\Delta_2}(\sigma_2)$ and $\phi_{B_2}(\sigma_2) \neq 0$ for almost all σ_2. Hence, $\phi_{\Delta_2}(\sigma_2)$ is identified. Similarly, ϕ_{Δ_1} is identified.

Finally, if Assumption 2.6 holds,

$$\begin{aligned}\phi_V(\sigma_1, \sigma_2) &= \int \exp(i(\sigma_1(b_1 + \delta_1) + \sigma_2 b_2)) f_B(b) f_{\Delta_1}(\delta_1) db d\delta_1 \\ &= \int \exp(i(\sigma_1 b_1 + \sigma_2 b_2)) f_B(b) db \int \exp(i\sigma_1\delta_1) f_{\Delta_1}(\delta_1) d\delta_1 = \phi_B(\sigma)\phi_{\Delta_1}(\sigma_1).\end{aligned} \tag{A.12}$$

As ϕ_V and ϕ_{Δ_1} are already identified, ϕ_B is identified. The proof of part (ii) of the theorem now follows from Fourier inversion applied to ϕ_B, ϕ_{Δ_1}, and ϕ_{Δ_2}. □

Proof of Theorem 2.2: In what follows, we write $U_j = (U_{0,j}, U_{-0,j})$, where $U_{0,j} = B_{0,j} + \check{B}_j\check{X}_j$, and $U_{-0,j} = B_{-0,j}$ is the random coefficients on the non-constant components of X_j. Similarly, we write $V_j = (V_{0,j}, V_{-0,j})$ with $V_{0,j} = \Delta_{0,j} + \check{\Delta}_j\check{X}_j$ and $V_{-0,j} = \Delta_{-0,j}$.

The proof of the identification of the distributions of U and V is identical to that of Theorem 2.1 (i) except for the presence of $\check{X}$. Therefore, argue as in the proof of Theorem 2.1 (i) and apply Proposition 1 with $(\check{X}, W)$ as conditioning variables, which identifies $\phi_{U|\check{X},W}$ and $\phi_{U+V|\check{X},W}$, the characteristic functions of U and $U + V$ conditional on $(\check{X}, W)$. By Assumption 2.6, it follows that $\phi_{U+V|\check{X},W}(\sigma \mid w) = \phi_{U|\check{X},W}(\sigma \mid \check{x}, w)\phi_{V|\check{X},W}(\sigma \mid \check{x}, w)$ and $\phi_{V|\check{X},W}(\sigma \mid \check{x}, w)$ is well-defined almost everywhere. Deduce from this that $\phi_{V|\check{X},W}$ is identified.

(i) For each j, let $\phi_{U_j|\check{X},W}(\sigma \mid w)$ be the characteristic function of $U_j \mid \check{X}, W$. Then, by the previous step, the map $\sigma_1 \mapsto \phi_{U_{0,j}|\check{X},W}(\sigma_1 \mid \check{x}, w) = \phi_{U_j|\check{X},W}((\sigma_1, 0, \cdots, 0) \mid \check{x}, w)$ is identified. By a property of the characteristic function and (2.8), one has

$$\begin{aligned} i^{-1}\frac{\partial}{\partial\sigma_1}\phi_{U_{0,j}|\check{X},W}(\sigma_1 \mid \check{x}, w)\Big|_{\sigma_1=0} &= \mathbb{E}[U_{0,j} \mid \check{X} = \check{x}, W = w] \\ &= \mathbb{E}[B_{0,j} + \check{B}_j\check{X}_j | \check{X} = \check{x}, W = w] = \mathbb{E}[B_{0,j} \mid W = w] + \mathbb{E}[\check{B}_j \mid W = w]\check{x}_j, \end{aligned} \tag{A.13}$$

where the last equality follows from the mean independence assumption. Taking expectations with respect to $\check{X}$, one then obtains,

$$\mathbb{E}\Big[i^{-1}\frac{\partial}{\partial\sigma_1}\phi_{U_{0,j}|W}(\sigma_1 \mid \check{X}, W)\Big|_{\sigma_1=0}(1, \check{X}_j') \mid W\Big] = \mathbb{E}[(B_j, \check{B}_j) \mid W]\mathbb{E}[(1, \check{X}_j')'(1, \check{X}_j') \mid W]. \tag{A.14}$$

By hypothesis, $\mathbb{E}[(1, \check{X}_j')'(1, \check{X}_j')|W]$ is positive definite and is hence invertible. Therefore,

$\mathbb{E}[(B_{0,j}, \check{B}_j)|W]$ is identified. The proof of identification of $\mathbb{E}[(\Delta_{0,j}, \check{\Delta}_j)|W]$ is analogous and therefore omitted.

(ii) By the previous step, $\phi_{(U_{0,1},U_{0,2})|\check{X},W}$ is identified. By assumption, W_j is a binary variable for $j = 1, 2$ and hence (2.8) implies the following objects are identified:

$$\phi_{(U_{0,1},U_{0,2})|\check{X},W}(\sigma_1, \sigma_2 \mid (0,0), w) = \phi_{(B_{0,1},B_{0,2})|W}(\sigma_1, \sigma_2 \mid w) \tag{A.15}$$

$$\phi_{(U_{0,1},U_{0,2})|\check{X},W}(\sigma_1, \sigma_2 \mid (1,1), w) = \phi_{(B_{0,1}+\check{B}_1,B_{0,2}+\check{B}_2)|W}(\sigma_1, \sigma_2 \mid w), \tag{A.16}$$

where $(\sigma_1, \sigma_2) \in \mathbb{R}^2$, and the equalities follow from $U_j = B_{0,j} + \check{B}_j\check{X}_j$ and $(B, \check{B}) \perp \check{X} \mid W$. By the conditional independence assumption $(B_{0,1}, B_{0,2}) \perp \check{B} \mid W$, it follows that

$$\phi_{(B_{0,1}+\check{B}_1,B_{0,2}+\check{B}_2)|W}(\sigma_1, \sigma_2 \mid w) = \phi_{(B_{0,1},B_{0,2})|W}(\sigma_1, \sigma_2 \mid w)\phi_{(\check{B}_1,\check{B}_2)|W}(\sigma_1, \sigma_2 \mid w). \tag{A.17}$$

By assumption, $\phi_{(B_{0,1},B_{0,2})|W}$ is nonzero almost everywhere. Therefore, $\phi_{(\check{B}_1,\check{B}_2)|W}(\sigma_1, \sigma_2 \mid w)$ is identified for almost all (σ_1, σ_2). The conclusion then follows from Fourier inversion. □

Proof of Theorem 2.3: We show the result with W null below. By Lemma A.2 the observations in the model are equivalent to projections of f_B on two dimensional subspaces of $\mathbb{R}^{2k}$ defined by $z = bx$. Here $z = (z_1, z_2)'$ parametrizes this plane. We denote by $\mathcal{P}_x$ the set of all observed projections, i.e $\mathcal{P}_x = \{(z_1x_1', z_2x_2')'|(z_1', z_2')' \in S_z\}$. If there is a non-trivial homogeneous polynomial $p : \mathbb{R}^{2k} \to \mathbb{R}$ that vanishes on

$$\bigcap_{x\in S_X} \mathcal{P}_x,$$

a second random variable $\tilde{B}$ exists which does not equal B but is observationally equivalent to B. This statement was proved by an explicit construction for $\tilde{B}$ in Cuesta-Alberto, et al. (2007) Theorem 3.5.

Without loss of generality we assume that $X_{i,j} = X_{1,1}$ and we split $a'\tilde{X} = a_1'\tilde{X}_1 + a_2'\tilde{X}_2$ with $a_1 \in \mathbb{R}^{k-1}$, $a_2 \in \mathbb{R}^k$, $\tilde{X}_1 = (X_{0,1}, X_{2,1}, X_{3,1}, \ldots, X_{k,1})'$. and $\tilde{X}_2 = (X_{0,2}, X_{1,2}, \ldots, X_{k,2})'$. By assumption, neither a_1 or a_2 is identically 0. Define the homogeneous polynomial of degree 2

$$p(X) \equiv X_{0,2}(X_{1,1} - a_1'\tilde{X}_1) - X_{0,1}a_2'\tilde{X}_2.$$

Then for every $x \in S_X$ and for every $y \in \mathcal{P}_x$

$$p(y) = z_2(z_1x_{1,1} - z_1a_1'\tilde{x}_1) - z_1z_2a_2'\tilde{x}_2 = z_1z_2(x_{1,1} - a_1'\tilde{x}_1 -_2' \tilde{x}_2) = 0.$$

Hence, p vanishes on $\bigcap_{x\in S_X} \mathcal{P}_x$. This implies that the distribution of B is not identified. □

Proof of Corollary 2.1: (i) Note that by Lemma A.3

$$\int_{S_{Z|X}} \exp(itz)\psi_{00}(z,x)dz = \|x\| \int_{\mathbb{R}^k\times\mathbb{R}^k} \exp(itbx) f_B(b)db\,. \tag{A.18}$$

is identified. The integral on the right hand side reads

$$\begin{aligned}\int_{\mathbb{R}^k\times\mathbb{R}^k} \exp(itbx) f_B(b)db &= \int_{\mathbb{R}^k\times\mathbb{R}^k} \exp\left(it \sum_{r=0,s=1}^{k-1,2} b_{r,s}x_{r,s}\right) f_B(b)db \\ &= \int_{\mathbb{R}^k\times\mathbb{R}^k} \exp\left(it \sum_{r\neq i,s\neq j} (b_{r,s}+a_{r,s}b_{i,j})x_{r,s}\right) f_B(b)db. \end{aligned}\tag{A.19}$$

Now the identification of the joint distribution of the coefficients in (2.12) follows by the argument in the proof of Theorem 2.1. Applying the same argument to ψ_{11} and replacing $b_{r,s}$ with $b_{r,s}+\Delta_{r,s}$ yields identification of the joint distribution of the coefficients $B_{r,s}+\Delta_{r,s}+a_{r,s}(B_{i,j}+\Delta_{i,j})$ with $(r,s)\neq(i,j)$. Apply the deconvolution argument as in the proof of Theorem 2.1. Then, the identification of the joint distribution of the coefficients in (2.13) follows. This establishes (i).

Without loss of generality, suppose $j=1$. Mimic the argument above and apply it to ψ_{10}. Then, it yields

$$\int_{S_{Z|X}} \exp(itz)\psi_{10}(z,x)dz = \|x\| \int_{\mathbb{R}^k\times\mathbb{R}^k} \exp\left(it \sum_{r\neq i,s\neq j} (u_{r,s}+a_{r,s}u_{i,j})x_{r,s}\right) f_U(u)du, \tag{A.20}$$

where $U=(B_1, B_2+\Delta_2)$. By the argument in the proof of Theorem 2.1, the joint distribution of the following coefficients are identified:

$$\begin{aligned}\Big(&B_{0,1}+a_{0,1}B_{i,1},\cdots, B_{i-1,1}+a_{i-1,1}B_{i,1}, B_{i+1,1}+a_{i+1,1}B_{i,1},\cdots \\ &B_{0,2}+\Delta_{0,2}+a_{0,2}(B_{i,1}+\Delta_{i,1}),\cdots, B_{k-1,2}+\Delta_{k-1,2}+a_{k-1,2}(B_{i,1}+\Delta_{i,1})\Big).\end{aligned}\tag{A.21}$$

Applying the argument to ψ_{01} identifies the joint distribution of following coefficients:

$$\begin{aligned}\Big(&B_{0,1}+\Delta_{0,1}+a_{0,1}(B_{i,1}+\Delta_{i,1}),\cdots, B_{i-1,1}+\Delta_{i-1,1}+a_{i-1,1}(B_{i,1}+\Delta_{i,1}), \\ &B_{i+1,1}+\Delta_{i+1,1}+a_{i+1,1}(B_{i,1}+\Delta_{i,1}),\cdots, \\ &B_{0,2}+a_{0,2}B_{i,1},\cdots, B_{k-1,2}+a_{k-1,2}B_{i,1}\Big).\end{aligned}\tag{A.22}$$

By the deconvolution argument as in the proof of Theorem 2.1, the joint distribution of the coefficients in (2.15) and that of (2.16) are identified. Finally, apply the deconvolution argument in Theorem (2.1) again to deconvolve the distribution of the coefficients in (2.14) from those in (A.21) and that of (2.16), which we identified in the previous step. This establishes (ii). □

The Definitions of the Sets in Corollary 2.2:

For each $(y_1, y_2) \in \{0,1\}^2$, let $R_1(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c)$ denote the set of values for $\Theta = (B_{1,1}, B_{1,2}, \Delta_{1,1}, \Delta_{1,2})$ under which the model predicts (y_1, y_2) as the unique PSNE under Assumption 2.2. Similarly, for each $(y_1, y_2) \in \{0,1\}^2$, let $R_2(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c)$ be the set of values $(b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2})$ under which (y_1, y_2) is predicted as one of multiple PSNEs under Assumption 2.2. Following an argument similar to Bresnahan & Reiss (1990,1991) and Tamer (2003), we have:

$$\begin{aligned}
R_1(0,0|\tilde{b}, \tilde{\delta}, x^c, z^c) &= \Big\{(b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2}) \in \mathbb{R}^4 | z_j^c < b_{1,j}x_{1,j}^c + \tilde{b}_j \tilde{x}_j^c, j = 1,2\Big\}, && \text{(A.23)}\\
R_1(0,1|\tilde{b}, \tilde{\delta}, x^c, z^c) &= \Big\{(b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2}) \in \mathbb{R}^4 | z_1^c < b_{1,1}x_{1,1}^c + \tilde{b}_1 \tilde{x}_1^c, z_2^c > b_{1,2}x_{1,2}^c + \tilde{b}_2 \tilde{x}_2^c\Big\}\\
&\quad \cup \Big\{b_{1,1}x_{1,1}^c + \tilde{b}_1\tilde{x}_1^c < z_1^c < (b_{1,1} + \delta_{1,1})x_{1,1}^c + (\tilde{b}_1 + \tilde{\delta}_1)\tilde{x}_1^c,\\
&\quad z_2^c > (b_{1,2} + \delta_{1,2})x_{1,2}^c + (\tilde{b}_2 + \tilde{\delta}_2)\tilde{x}_2^c\Big\}, && \text{(A.24)}\\
R_1(1,0|\tilde{b}, \tilde{\delta}, x^c, z^c) &= \Big\{(b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2}) \in \mathbb{R}^4 | z_1^c > (b_{1,1} + \delta_{1,1})x_{1,1}^c + (\tilde{b}_1 + \tilde{\delta}_1)\tilde{x}_1^c,\\
&\quad z_2^c > (b_{1,2} + \delta_{1,2})x_{1,2}^c + (\tilde{b}_2 + \tilde{\delta}_2)\tilde{x}_2^c\Big\}\\
&\quad \cup \Big\{b_{1,1}x_{1,1}^c + \tilde{b}_1\tilde{x}_1^c < z_1^c < (b_{1,1} + \delta_{1,1})x_{1,1}^c + (\tilde{b}_1 + \tilde{\delta}_1)\tilde{x}_1^c,\\
&\quad z_2^c < b_{1,2}x_{1,2}^c + \tilde{b}_2\tilde{x}_2^c\Big\}, && \text{(A.25)}\\
R_1(1,1|\tilde{b}, \tilde{\delta}, x^c, z^c) &= \Big\{(b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2}) \in \mathbb{R}^4 | z_j^c > (b_{1,j} + \delta_{1,j})x_{1,j}^c + (\tilde{b}_j + \tilde{\delta}_j)\tilde{x}_j^c, j = 1,2\Big\}; && \text{(A.26)}
\end{aligned}$$

and

$$\begin{aligned}
&R_2(0,0|\tilde{b}, \tilde{\delta}, x^c, z^c) = R_2(1,1|\tilde{b}, \tilde{\delta}, x^c, z^c) = \emptyset && \text{(A.27)}\\
&R_2(0,1|\tilde{b}, \tilde{\delta}, x^c, z^c) = R_2(1,0|\tilde{b}, \tilde{\delta}, x^c, z^c)\\
&= \Big\{(b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2}) \in \mathbb{R}^4 | b_{1,j}x_{1,j}^c + \tilde{b}_j\tilde{x}_j^c < z_j^c < (b_{1,j} + \delta_{1,j})x_{1,j}^c + (\tilde{b}_j + \tilde{\delta}_j)\tilde{x}_j^c, j = 1,2\Big\}. && \text{(A.28)}
\end{aligned}$$

For each $j \in \{1,2\}$, let $\delta_{-1,j}$ denote a vector that stacks all components of δ_j except $\delta_{1,j}$. For each $(y_1, y_2) \in \{0,1\}^2$, let $R_1(y_1, y_2|\tilde{b}, x^c, z^c)$ be the set of values $(b_{1,1}, b_{1,2}, \delta)$ under which the model predicts (y_1, y_2) as the unique PSNE under Assumption 2.3. Similarly, let $R_2(y_1, y_2|\tilde{b}, \tilde{\delta}, x^c, z^c)$ be the set of values under which (y_1, y_2) is predicted as one of multiple PSNEs. These sets are given as

$$
\begin{aligned}
&R_1(0,0|\tilde{b},x^c,z^c)\\
&\quad=\Big\{(b_{1,1},b_{1,2},\delta)\in\mathbb{R}^{2+2k}|z_1^c>b_{1,1}x_{1,1}^c+\tilde{b}_1\tilde{x}_1^c, z_2^c>(b_{1,2}+\delta_{1,2})x_{1,2}^c+(\tilde{b}_2+\delta_{-1,2})\tilde{x}_2^c\Big\}\\
&\qquad\quad\cup\Big\{(b_{1,1}+\delta_{1,1})x_{1,1}^c+(\tilde{b}_1+\delta_{-1,1})\tilde{x}_1^c<z_1^c<b_{1,1}x_{1,1}^c+\tilde{b}_1\tilde{x}_1, z_2^c>b_{1,2}x_{1,2}^c+\tilde{b}_2\tilde{x}_2\Big\}\\
&R_1(0,1|\tilde{b},x^c,z^c)\\
&\quad=\Big\{(b_{1,1},b_{1,2},\delta)\in\mathbb{R}^{2+2k}|z_1^c>(b_{1,1}+\delta_{1,1})x_{1,1}^c+(\tilde{b}_1+\delta_{-1,1})\tilde{x}_1^c, z_2^c>b_{1,2}x_{1,2}^c+\tilde{b}_2\tilde{x}_2\Big\}\\
&R_1(1,0|\tilde{b},x^c,z^c)\\
&\quad=\Big\{(b_{1,1},b_{1,2},\delta)\in\mathbb{R}^{2+2k}|z_1^c>b_{1,1}x_{1,1}^c+\tilde{b}_1\tilde{x}_1^c, z_2^c<(b_{1,2}+\delta_{1,2})x_{1,2}^c+(\tilde{b}_2+\delta_{-1,2})\tilde{x}_2\Big\}\\
&R_1(1,1|\tilde{b},x^c,z^c)\\
&\quad=\Big\{(b_{1,1},b_{1,2},\delta)\in\mathbb{R}^{2+2k}|z_1^c<(b_{1,1}+\delta_{1,1})x_{1,1}^c+(\tilde{b}_1+\delta_{-1,1})\tilde{x}_1^c, z_2^c<b_{1,2}x_{1,2}^c+\tilde{b}_2\tilde{x}_2\Big\}\\
&\qquad\quad\cup\Big\{(b_{1,1}+\delta_{1,1})x_{1,1}^c+(\tilde{b}_1+\delta_{-1,1})\tilde{x}_1^c<z_1^c<b_{1,1}x_{1,1}^c+\tilde{b}_1\tilde{x}_1,\\
&\qquad\qquad\qquad\qquad\qquad z_2^c<(b_{1,2}+\delta_{1,2})x_{1,2}^c+(\tilde{b}_2+\delta_{-1,2})\tilde{x}_2^c\Big\};
\end{aligned}
\tag{A.29}
$$

and

$$
R_2(0,1|\tilde{b},x^c,z^c)=R_2(1,0|\tilde{b},x^c,z^c)=\emptyset \tag{A.30}
$$

$$
\begin{aligned}
&R_2(0,0|\tilde{b},x^c,z^c)=R_2(1,1|\tilde{b},x^c,z^c)\\
&=\Big\{(b_{1,1},b_{1,2},\delta)\in\mathbb{R}^{2+2k}|(b_{1,j}+\delta_{1,j})x_{1,j}^c+(\tilde{b}_j+\delta_{-1,j})\tilde{x}_j^c<z_j^c<b_{1,j}x_{1,j}^c+\tilde{b}_j\tilde{x}_j^c, j=1,2\Big\}.
\end{aligned}
\tag{A.31}
$$

Let $\theta=(b_{1,1},b_{1,2},\delta_{1,1},\delta_{1,2})$. For each $(\tilde{b},\tilde{\delta},w^c)$, the identified set $\mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}$ for the conditional density of θ in Corollary 2.2 (i) is defined as

$$
\begin{aligned}
\mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}=\Big\{f\in L^1(\mathbb{R}^4)|f\geq 0,\ &\int f(\theta\mid\tilde{b},\tilde{\delta},w^c)d\theta=1,\\
&\int e^{it_1(b_{1,1}+b_{1,2})+it_2(\delta_{1,1}+\delta_{1,2})}f(\theta\mid\tilde{b},\tilde{\delta},w^c)d\theta\\
&=\phi_{B_{1,1}+B_{1,2}|\tilde{B},W}(t_1\mid\tilde{b},w^c)\phi_{\Delta_{1,1}+\Delta_{1,2}|\tilde{\Delta},W}(t_2\mid\tilde{\delta},w^c),\ \forall(t_1,t_2)\in\mathbb{R}^2\Big\}.
\end{aligned}
\tag{A.32}
$$

We note that $\phi_{B_{1,1}+B_{1,2}|\tilde{B},W}$ and $\phi_{\Delta_{1,1}+\Delta_{1,2}|\tilde{\Delta},W}$ are identified by Corollary 2.1. Hence, all restrictions on f are linear. The same comment applies to $\mathcal{F}_{I,\tilde{b},w^c}$ defined below.

Let $\theta=(b_{1,1},b_{1,2},\delta)$. For each $(\tilde{b},w^c)$, the identified set $\mathcal{F}_{I,\tilde{b},w^c}$ for the conditional density of

θ in Corollary 2.2 (ii) is defined as follows:

$$\mathcal{F}_{I,\tilde{b},w^c} = \Big\{ f \in L^1(\mathbb{R}^{2+2k}) | f \geq 0, \quad \int f(\theta \mid \tilde{b}, w^c) d\theta = 1,$$
$$\int e^{it_1(b_{1,1}+b_{1,2})+it_2'\tilde{\delta}_1} f(\theta \mid \tilde{b}, w^c) d\theta = \phi_{B_{1,1}+B_{1,2}|\tilde{B},W}(t_1 \mid \tilde{b}, w^c)\phi_{\tilde{\Delta}_1|W}(t_2 \mid w^c), \ \forall t_1 \in \mathbb{R}, t_2 \in \mathbb{R}^{k-1}$$
$$\int e^{it_1(b_{1,1}+b_{1,2})+it_2(\delta_{1,1}+\delta_{1,2})+it_3'\tilde{\delta}_2} f(\theta \mid \tilde{b}, w^c) d\theta = \phi_{B_{1,1}+B_{1,2}|\tilde{B}}(t_1 \mid \tilde{b}, w^c)\phi_{\Delta_{1,1}+\Delta_{1,2},\tilde{\Delta}_2|W}(t_2, t_3 \mid w^c),$$
$$\forall (t_1, t_2) \in \mathbb{R}^2, t_3 \in \mathbb{R}^{k-1} \Big\}. \tag{A.33}$$

□

Proof of Corollary 2.2: (i) The counterfactual probability for $(y_1, y_2) \in \{0,1\}^2$ is

$$\mathbb{P}(y_1, y_2 | x^c, z^c, w^c) = \int \mathbb{P}(y_1, y_2 | \tilde{b}, \tilde{\delta}, x^c, z^c, w^c) f_{\tilde{B}|W}(\tilde{b} \mid w^c) f_{\tilde{\Delta}|W}(\tilde{\delta} \mid w^c) d\tilde{b} d\tilde{\delta}, \tag{A.34}$$

where $\mathbb{P}(y_1, y_2 | \tilde{b}, \tilde{\delta}, x^c, z^c, w^c)$ is the conditional probability of $(Y_1, Y_2) = (y_1, y_2)$ given $(\tilde{B}, \tilde{\Delta}, X, Z, W) = (\tilde{b}, \tilde{\delta}, x^c, z^c, w^c)$, and we note that $f_{\tilde{B}|W} f_{\tilde{\Delta}|W}$ is identified by Corollary 2.1. By the definitions of R_1 and R_2 and letting $\theta = (b_{1,1}, b_{1,2}, \delta_{1,1}, \delta_{1,2})$, we have

$$\begin{aligned} \int_{R_1(y_1,y_2|\tilde{b},\tilde{\delta},x^c,z^c)} & f_{B_{1,1},B_{1,2},\Delta_{1,1},\Delta_{1,2}|\tilde{B},\tilde{\Delta},W}(\theta | \tilde{b}, \tilde{\delta}, w^c) d\theta \\ & \leq \mathbb{P}(y_1, y_2 | \tilde{b}, \tilde{\delta}, x^c, z^c, w^c) \\ & \leq \int_{R_1(y_1,y_2|\tilde{b},\tilde{\delta},x^c,z^c) \cup R_2(y_1,y_2|\tilde{b},\tilde{\delta},x^c,z^c)} f_{B_{1,1},B_{1,2},\Delta_{1,1},\Delta_{1,2}|\tilde{B},\tilde{\Delta},W}(\theta | \tilde{b}, \tilde{\delta}, w^c) d\theta, \end{aligned} \tag{A.35}$$

for any $(\tilde{b}, \tilde{\delta}) \in S_{\tilde{B}|W} \times S_{\tilde{\Delta}|W}$, where $f_{B_{1,1},B_{1,2},\Delta_{1,1},\Delta_{1,2}|\tilde{B},\tilde{\Delta},W}$ is the conditional density function of $(B_{1,1}, B_{1,2}, \Delta_{1,1}, \Delta_{1,2})$ given $(\tilde{B}, \tilde{\Delta}, W)$. This density is not identified but belongs to the identified set $\mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}$ in (A.32) due to Corollary 2.1 (i) and Assumption 2.6. Take an infimum (and supremum) on the left hand side (and the right hand side) of (A.35) with respect to the conditional density over $\mathcal{F}_{I,\tilde{b},\tilde{\delta},w^c}$. Then,

$$\mathbb{P}_L(y_1, y_2 | \tilde{b}, \tilde{\delta}, x^c, z^c, w^c) \leq \mathbb{P}(y_1, y_2 | \tilde{b}, \tilde{\delta}, x^c, z^c, w^c) \leq \mathbb{P}_U(y_1, y_2 | \tilde{b}, \tilde{\delta}, x^c, z^c, w^c). \tag{A.36}$$

Multiplying $f_{\tilde{B}|W} f_{\tilde{\Delta}|W}$, which is identified by Corollary 2.1, to these bounds and integrating them yields the conclusion of (i).

(ii) Write the counterfactual probability for $(y_1, y_2) \in \{0,1\}^2$ as

$$\mathbb{P}(y_1, y_2 | x^c, z^c, w^c) = \int \mathbb{P}(y_1, y_2 | \tilde{b}, x^c, z^c, w^c) f_{\tilde{B}|W}(\tilde{b} \mid w^c) d\tilde{b}, \tag{A.37}$$

where $\mathbb{P}(y_1, y_2 | \tilde{b}, x^c, z^c, w^c)$ is the conditional probability of $(Y_1, Y_2) = (y_1, y_2)$ given $(\tilde{B}, X, Z, W) =$

$(\tilde{b}, x^c, z^c, w^c)$. By the definitions of R_1 and R_2 and letting $\theta = (b_{1,1}, b_{1,2}, \delta)$, we have

$$\begin{aligned}
&\int_{R_1(y_1,y_2|\tilde{b},x^c,z^c)} f_{B_{1,1},B_{1,2},\Delta|\tilde{B},W}(\theta|\tilde{b}, w^c)d\theta \\
&\qquad\qquad\qquad \leq \mathbb{P}(y_1, y_2|\tilde{b}, x^c, z^c, w^c) \\
&\qquad\qquad\qquad \leq \int_{R_1(y_1,y_2|\tilde{b},x^c,z^c)\cup R_2(y_1,y_2|\tilde{b},x^c,z^c)} f_{B_{1,1},B_{1,2},\Delta|\tilde{B},W}(\theta|\tilde{b}, w^c)d\theta, \quad \text{(A.38)}
\end{aligned}$$

for any $\tilde{b} \in S_{\tilde{B}|W}$. Note that the conditional density $f_{B_{1,1},B_{1,2},\Delta|\tilde{B},W}$ belongs to the identified set $\mathcal{F}_{I,\tilde{b},w^c}$ by Corollary 2.1 (ii) and Assumption 2.6. The rest of the argument is then the same as (i).

$\square$